НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ХАРЧОВИХ ТЕХНОЛОГІЙ

РАДЗІЄВСЬКАО.І.
КУЗЬМІНСЬКА Н.Л.
ВАСЮТИНСЬКА Ю.О.

СПЕЦІАЛЬНІ РОЗДІЛИ МАТЕМАТИКИ ДЛЯ ЕКОНОМІСТІВ

(частина І, ІІ)

Рекомендовано Вченою радою

Національного університету харчових технологій

як навчальний посібник

для студентів вищих навчальних закладів

Київ, НУХТ, 2016

УДК 519.21/519.22

Рецензенти: д-р фіз.-мат. наук, проф. Зелінський Ю.Б. (Інститут математики НАНУ); д-р фіз.-мат. наук, Романюк А.С. **(**Інститут математики НАНУ); д-р екон. наук, доц., Пєтухова О.М. (Національний університет харчових технологій).

Спеціальні розділи математики для економістів (ч. I, II) : навчальний посібник/ О.І. Радзієвська, Н.Л. Кузьмінська, Ю.О. Васютинська. – К. : НУХТ, 2016. – 324 с.

Навчальний посібник написано згідно з програмою курсу «Теорія ймовірностей та математична статистика» для студентів економічних спеціальностей. Лекції попри теоретичний матеріал містять багато прикладів, розв’язання яких приведене не тільки аналітично, але і за допомогою сучасних інформаційних технологій (програм MS Excel, MathCad). Практичні заняття побудовано на спеціально підібраних задачах, які мають економічну спрямованість. Приведено зразки тестів та контрольних робіт. Рекомендовано студентам економічних спеціальностей очної і заочної форм навчання, аспірантам, викладачам та фахівцям, що вивчають дисципліну самостійно.

ЗМІСТ

ВСТУП

Теорія ймовірностей є математичною наукою, яка вивчає кількісні закономірності випадкових масових явищ, а також є теоретичною основою математичної та прикладної статистики. У зв'язку з розвитком масових процесів у виробництві і в економіці, а також через необхідність більш тонкого аналізу результатів експерименту, математичного моделювання і розв'язання оптимізаційних задач, ймовірнісні і статистичні методи широко проникли в економічні, технічні та технологічні науки.

Курс «Теорія ймовірностей та математична статистика» складає основну частину необхідної навчальної бази при вивченні таких фахових економічних дисциплін як економетрика, багатомірні статистичні методи, економічний аналіз, теорія ризиків та ін. Велика кількість літератури з даної дисципліни забезпечує сучасному студенту можливість вибору доступного та різнорівневого її вивчення. Але, в свою чергу, це висуває високі вимоги до написання якісного підручника, який би полегшив вивчення обраної дисципліни.

Навчальний посібник «Спеціальні розділи математики для економістів» складається з лекційних і практичних занять, містить зразки тестів та контрольних робіт, предметний покажчик. Виклад теоретичного матеріалу у вигляді лекцій є максимально доступним і супроводжується розв'язанням прикладів та задач. Кожну лекцію підручника завершує практичне заняття, у якому зібрано великий набір завдань різноманітної складності для самостійного опрацювання. Зразки тестів та контрольних робіт сприятимуть організації самоконтролю вивченого матеріалу. Предметний покажчик стане корисним для прискорення пошуку потрібної інформації у посібнику.

Засвоєння курсу є важливим як в аспекті розуміння місця основ теорії ймовірностей та математичної статистики в економічних дослідженнях, так і в оволодінні методикою та технікою статистичних обчислень з використанням сучасних комп'ютерних технологій. Тому особливістю даного посібника є поєднання зразків стандартного розв’язання завдань з розв’язання їх, користовуючись інформаційними технологіями.

Для розв’язування задач пропонується використання програми *MS Excel*, яка є потужним програмним засобом для роботи з таблицями даних, що дозволяє обчислювати значення, упорядковувати, аналізувати і графічно представляти різні види даних, та системи *MathCad*, яка є сучасним програмним продуктом, який дозволяє виконувати різні наукові та інженерні розрахунки, починаючи від елементарної математики і закінчуючи складними реалізаціями чисельних методів.

Навчальний посібник «Спеціальні розділи математики для економістів» рекомендовано до використання студентам економічних спеціальностей очної і заочної форм навчання, аспірантам, викладачам та фахівцям, що вивчають дисципліну самостійно.

Частина I. Теорія ймовірностей

Лекція 1. Основні поняття теорії ймовірностей. Елементи комбінаторики

1.1. Предмет теорії ймовірностей та математичної статистики

1.2. Випадкові події. Алгебра подій

1.3. Означення ймовірності

1.4. Елементи комбінаторики

1.1. Предмет теорії ймовірностей та математичної статистики

Теорія ймовірностей та математична статистика вивчає закономірності масових випадкових явищ. Методи теорії ймовірностей та математичної статистики широко застосовуються у різних галузях науки і техніки: теорії масового обслуговування, теорії інформації, теорії похибок спостереження, аналізі технологічних процесів, контролі якості продукції, економетричному моделюванні тощо.

Теорія ймовірностей виникла приблизно у середині XVII століття. Її перші кроки пов’язані з іменами Б. Паскаля, П. Ферма, Я. Бернуллі.

Спочатку до теорії ймовірностей відносили задачі, пов’язані із різними азартними іграми (карти, кості тощо). У цей період її апарат зводився суто до арифметичних і комбінаторних прийомів.

Вимоги з боку природничих та соціальних наук, техніки привели до необхідності подальшого розвитку теорії ймовірностей і застосування більш досконалого аналітичного апарату.

Особливого значення у розвитку аналітичних методів теорії ймовірностей мали праці А. Муавра, П.-С. Лапласа, К.Ф. Гаусса С.Д. Пуассона (рубіж XVIII та XIX ст.).

Із середини XIX ст. і на початку XX ст. розвиток теорії ймовірностей пов'язаний із іменами П.Л. Чебишова, А.М. Ляпунова, А.А. Маркова. У XX ст.

видатний вклад внесли роботи С.Н. Бернштейна, Б.В. Гнеденка, А.Н. Колмагорова, П. Леві, А.В. Скорохода, Є.Є. Слуцького, В. Феллера, А.Я. Хінчина та ін.

1.2. Випадкові події. Алгебра подій

Кожна наука, яка розробляє загальну теорію деякого кола явищ, містить низку основних понять, на яких вона базується. Такими, наприклад, у геометрії є поняття точки, прямої, площини, у механіці – поняття сили, маси, швидкості тощо. Зрозуміло, що не всі основні поняття можуть бути строго визначені, оскільки визначити поняття – це означає звести його до інших, які вже існують. Очевидно, що процес визначення одних понять через інші повинен на якомусь етапі завершуватись.

У теорії ймовірностей до основних понять відносяться: стохастичні випробування, подія, ймовірність події.

Означення. Стохастичними випробуваннями називаються випробування, результати яких не можна наперед передбачити.

Підкидається шестигранний гральний кубик, грані якого пронумеровані від 1 до 6. На перед не можна передбачити скільки очок випаде на його грані.

Означення. Подією називається будь-яке явище, відносно якого можна сказати, відбулося воно чи не відбулося.

Усі події можна поділити на випадкові, достовірні та неможливі.

Означення. Випадковою подією називається наслідок стохастичних випробувань. Вони позначаються великими літерами латинського алфавіту, наприклад, A, B, C.

Поява грані з номером 6 при підкиданні грального кубика – випадкова подія; поява герба при підкиданні монети – теж випадкова подія.

Означення. Достовірною подією називається подія, яка обов'язково відбудеться при виконанні певного комплексу умов. Достовірна подія позначається великою літерою грецького алфавіту Ω (омега).

Якщо в урні лежать білі кулі, то подія, яка полягає у тому, що з урни витягли білу кулю – достовірна подія.

<u>*Означення.*</u> *Неможливою подією* називається подія, яка не може відбутися при виконанні певного комплексу умов. Неможливу подію позначається, як порожня множина – $\varnothing$.

Якщо в урні лежать тільки білі кулі, то подія, яка полягає у тому, що з урни витягли чорну кулю – неможлива подія.

Події, які є наслідками стохастичних випробувань можуть бути пов'язані між собою певними співвідношеннями, які будуть визначені далі.

<u>*Означення.*</u> Події A і B називаються *несумісними*, якщо поява однієї з них виключає появу іншої в одному і тому ж випробуванні.

Поява грані з 6 очками при одному підкиданні виключає появу грані з 3 очками.

<u>*Означення.*</u> *Протилежною подією* до події A називається така подія, яка є несумісною з подією A, і обов'язково одна з цих двох подій настане у результаті випробування. Протилежна подія позначається через $\overline{A}$.

Стрілець робить один залп по мішені. Тоді, якщо подія A – стрілець влучив у мішень, то $\overline{A}$ – стрілець не влучив у мішень.

Зауваження. Події A і $\overline{A}$ несумісні, але несумісні події не обов'язково будуть протилежними.

Події, які полягають у тому, що на гральному кубику випаде 6 і 3 очки при одному підкиданні, є несумісними, але не є протилежними.

<u>*Означення.*</u> Події $B_1, B_2, ..., B_n$ утворюють *повну групу подій*, якщо вони попарно несумісні, а всі разом є достовірною подією.

З кожним випробуванням (стохастичним експериментом) пов'язана низка подій, які можуть з'являтись одночасно. Наприклад, нехай при підкиданні грального кубика подія A полягає у тому, що випало 3 очка, а подія B – непарне число очок. Зрозуміло, що ці події не виключають одна одну.

Нехай усі можливі наслідки випробування здійснюються в низці єдино можливих частинних випадків, які є взаємовиключними. Такі частинні випадки називаються *елементарними подіями*, а множину всіх елементарних подій називають простором елементарних подій і позначають літерою Ω. Тоді: 1) кожен наслідок випробування представлено одною і тільки одною елементарною подією; 2) будь-яка подія *A*, яка пов'язана з цим випробуванням, є множина (сукупність) скінченого або нескінченого числа елементарних подій; 3) подія *A* відбудеться тоді і тільки тоді, коли відбудеться одна з елементарних подій, що входять у цю множину.

Наприклад, нехай подія *A* полягає в тому, що при одному підкиданні грального кубика випало непарне число очок. Простір елементарних подій цього випробування – $\Omega = \{1; 2; 3; 4; 5; 6\}$, подія *A* буде складатись із елементарних подій $\{1; 3; 5\}$. Тому подію *A* можна розглядати як підмножину множини Ω.

Взагалі, будь-який наслідок експерименту можна вважати підмножиною простору елементарних подій. Отже, аналогічно теорії множин будується алгебра подій.

<u>*Означення.*</u> *Сумою двох подій A* і *B* називається подія $A + B = A \cup B$, яка має місце тоді і тільки тоді, коли відбулася хоча б одна з подій *A* або *B*.

У загальному випадку під сумою декількох подій розуміється подія, яка полягає у тому, що хоча б одна з них відбудеться.

Нехай подія А полягає у тому, що у мішень влучив перший стрілець, *B* – у мішень влучив другий стрілець. Тоді подія $A + B$ – влучення у мішень хоча б одним стрільцем, можливо, двома одразу.

<u>*Означення.*</u> *Добутком двох подій A* і *B* називається подія $AB = A \cap B$, яка полягає в одночасній появі як події *A*, так і події *B*.

У загальному випадку під добутком декількох подій розуміється подія, яка полягає у сумісній появі всіх цих подій.

Нехай подія A полягає у тому, що при одному підкиданні кубика випаде парне число очок, B – випаде число очок, що ділиться на 3. Тоді подія AB – випаде 6 очок.

Ці основні операції над подіями наочно ілюструються діаграмами Венна (рис. 1.1), на яких множина Ω є множиною точок квадрата, а події A і B – множинами точок кругів.

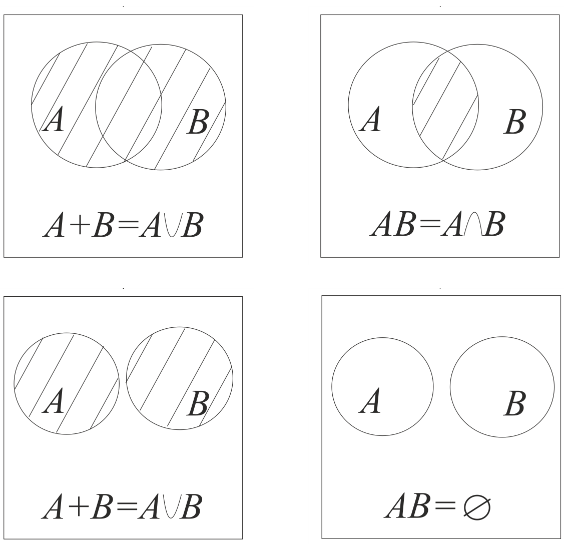


Рис. 1.1

Властивості операцій.

1. $A+B=B+A$; $AB=BA$ (комутативний закон).
2. $(A+B)+C=A+(B+C)$; $(AB)C=A(BC)$ (асоціативний закон).
3. $A+A=A.$
4. $A+B=A$, якщо $B\subset A.$
5. $AA=A$.
6. $AB=B.$
7. $A(B+C)=AB+AC.$
8. $AB+C=(A+C)(B+C).$
9. $A+\overline{A}=\Omega.$
10. $A\subset\overline{A}=\varnothing.$

Приклад 1.1. Стрілець двічі стріляє по мішені. Нехай подія A_1 полягає у тому, що він влучив при першому пострілі, A_2 – при другому. Описати простір елементарних подій. Записати події: 1) стрілець влучив у мішень тільки один раз (B_1); 2) стрілець влучив у мішень хоча б один раз (B_2); 3) стрілець не влучив у мішень жодного разу (B_3).

Розв'язання. Події $\overline{A}_1, \overline{A}_2$ полягають у тому, що при відповідному пострілі стрілець не влучив у мішень. Тоді простір елементарних подій складатиметься з наступних чотирьох подій $\Omega = \{A_1A_2; A_1\overline{A}_2; \overline{A}_1A_2; \overline{A}_1\overline{A}_2\}$.

1) стрілець влучив у мішень тільки один раз – означає, що він влучив при першому пострілі, а при другому – не влучив, або навпаки, тобто

$$B_1 = A_1\overline{A}_2 + \overline{A}_1A_2;$$

2) стрілець влучив у мішень хоча б один раз – означає, що у мішені або одна пробоїна, або дві:

$$B_2 = A_1 + A_2 = A_1\overline{A}_2 + \overline{A}_1A_2 + A_1A_2;$$

3) стрілець не влучив у мішень жодного разу – означає, що він промахнувся двічі:

$$B_3 = \overline{A}_1\overline{A}_2.$$

Відповідь: $B_1 = A_1\overline{A}_2 + \overline{A}_1A_2$; $B_2 = A_1 + A_2 = A_1\overline{A}_2 + \overline{A}_1A_2 + A_1A_2$; $B_3 = \overline{A}_1\overline{A}_2$.

1.3. Означення ймовірності

1.3.1. Класичне означення ймовірності

Нехай простір елементарних подій Ω складається з n елементарних подій $\{\omega_1, \omega_2, ..., \omega_n\}$. Припустимо, що всі події цієї множини рівно можливі, тобто немає підстав вважати, що деякі події з цієї множини превалюють, в сенсі появи, над іншими.

Означення. Ймовірністю $P(A)$ події A називається відношення числа m рівно можливих елементарних наслідків, які сприяють появі події A, до загального числа n усіх рівно можливих і єдино можливих наслідків даного випробування.

Отже, якщо m – число елементарних наслідків, що сприяють появі події A, а n – загальне число усіх елементарних наслідків у даному випробуванні і всі наслідки рівноможливі, то за означенням:

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

Наведене означення називається класичним.

З означення ймовірності випливають наступні її властивості:

1) Оскільки $0 \le m \le n$, то

$$0 \le P(A) \le 1.$$

Отже, ймовірність будь-якої події є невід'ємне число, що не перевищує одиницю.

2) Ймовірність неможливої події дорівнює нулю.

Дійсно, якщо подія A – неможлива, то число наслідків, що сприяють цій події $m = 0$ і

$$P(A) = \frac{0}{n} = 0.$$

3) Ймовірність достовірної події дорівнює одиниці.

Дійсно, якщо A – достовірна подія, то число наслідків, що сприяють цій події $m = n$ і

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

Зауваження. З означення ймовірності маємо, що рівно можливі елементарні події є рівно ймовірними, тобто мають одну і ту ж ймовірність.

Приклад 1.2. В урні знаходяться 4 білих і 7 чорних куль. Яка ймовірність дістати з урни білу, чорну, червону кулю?

Розв'язання. Нехай подія A полягає у тому, що з урни дістали білу кулю, B – чорну кулю, C – червону кулю. Тоді

$$P(A)=\frac{4}{11},\ P(B)=\frac{6}{11},\ P(C)=\frac{0}{11}=0.$$

Відповідь: ймовірність дістати з урни білу кулю дорівнює $\frac{4}{11}$, чорну кулю – $\frac{6}{11}$, червону кулю – 0.

Приклад 1.3. Підкидають два гральних кубика. Яка ймовірність того, що у сумі на гранях випаде 4 очка.

Розв'язання. Всього можливих наслідків даного експерименту $n=36$. Нехай подія A полягає у тому, що у сумі на гранях двох кубиків випаде 4 очка. Це можливо, якщо на першому кубику випаде 1 очко, на другому – 3, або навпаки, і, якщо на обох кубиках випаде по 2 очка. Отже, $m=3$. Тоді $P(A)=\frac{3}{36}=\frac{1}{12}$.

Відповідь: ймовірність того, що у сумі на гранях випаде 4 очка становить $\frac{1}{12}$.

1.3.2. Статистичне означення ймовірності

Слід зауважити, що застосування класичного означення ймовірності значно обмежене: на практиці часто зустрічаються випробування з нескінченою кількістю різних можливих наслідків. Крім того немає загальних методів, які б дозволили результати випробування представити у вигляді суми рівно можливих елементарних наслідків, навіть зі скінченим їх числом. У таких випадках використовується статистичне означення ймовірності.

Нехай виконуються n одноманітних випробувань одним з наслідків яких є подія A.

Означення. Відношення числа m появ події A до загального числа випробувань n називається *відносною частотою* події A.

Відносна частота події A при n випробуваннях позначається $W_n(A)$. З означення: $W_n(A) = \frac{m}{n}$.

Очевидно, що $0 \le W_n(A) \le 1$.

При однотипових масових випробуваннях у багатьох випадках спостерігається стійкість відносної частоти події, тобто при числі випробувань $n \to \infty$, відносна частота $W_n(A)$ події A коливається біля деякого постійного числа p. Це число p називається ймовірністю події A в статистичному сенсі.

Означення. *Статистичною ймовірністю* події A називається майже достовірна границя її відносної частоти при необмежено зростаючому числі випробувань

$$P(A) = \lim_{n\to\infty} W_n(A).$$

1.3.3. Геометрична ймовірність

Класичне означення ймовірності відноситься до випадку, коли простір елементарних подій дискретний і скінчений. Якщо простір елементарних подій неперервний, то на цьому просторі прийнято розглядати геометричну ймовірність. Нехай простором елементарних подій є множина точок деякої області G, яка має певний вимір, а події A сприяє лише її частина $g \subset G$, тоді ймовірність події A визначається наступним чином.

Означення. *Геометрична ймовірність* випадкової події A дорівнює відношенню міри області g до міри G, тобто $P(A) = \frac{\text{mes}\, g}{\text{mes}\, G}$.

Через mes позначається міра області. Це може бути, наприклад, довжина, площа, об'єм.

Наведене означення називається *геометричним*.

<u>Приклад 1.4.</u> (Задача про зустріч). Два студенти домовились зустрітись біля університету з 12 до 13 год. і чекати один одного не більше 20 хв. Яка ймовірність, що вони зустрінуться.

<u>Розв'язання.</u> Нехай x – час приходу першого студента, а y – другого. Зрозуміло, що $12 \le x \le 13$, $12 \le y \le 13$. Тоді усі можливі наслідки є множина G – множина точок квадрата зі стороною рівною 1 (рис. 1.2), $S(G) = 1$.

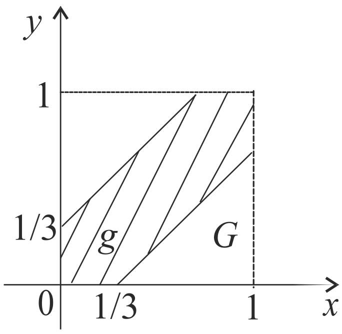


Рис. 1.2

За умовою студенти чекають один одного не більше 20 хв. або $\frac{1}{3}$ год., тобто $|y - x| \le \frac{1}{3}$ або $x - \frac{1}{3} \le y \le x + \frac{1}{3}$.

Множина точок g, що відповідає цій нерівності, і є наслідками випадкового експерименту, який сприяє зустрічі. Вона утворює шестикутник (рис 1.2) площею $S(g) = 1 - 2 \cdot \frac{1}{2} \cdot \frac{4}{9} = \frac{5}{9}$.

Отже, ймовірність події A (студенти зустрінуться):

$$P(A) = \frac{S(g)}{S(G)} = \frac{5}{9}.$$

<u>Відповідь:</u> ймовірність того, що студенти зустрінуться дорівнює $\frac{5}{9}$.

1.4. Елементи комбінаторики

Знаходження ймовірності випадкової події часто пов'язане з визначенням кількості можливих варіантів складання груп елементів із деякої множини, що відрізняються за складом елементів або за порядком їх вибору.

Для цього використовуються основні правила і формули комбінаторики – розділу математики, в якому вивчаються методи обчислення числа різноманітних виборів комбінацій елементів.

1.4.1. Основні правила комбінаторики

Правило додавання. Якщо об'єкт A_1 можна вибрати n_1 способами, об'єкт A_2 – іншими n_2 способами, A_3 – відмінними від перших двох n_3 способами і т.д., A_k – n_k способами, що відрізняються від перших $(k-1)$, то вибір одного з елементів або A_1, або A_2, …, або A_k можна здійснити $n_1 + n_2 + ... + n_k$ способами.

Приклад 1.5. У першій групі 5 відмінників, у другій – 6 відмінників. Скільки існує різних способів вибору одного відмінника із двох груп на олімпіаду з математики?

Розв'язання. Одного відмінника із першої групи можна вибрати 5-ма способами, з другої – 6-ма способами. Отже, із двох груп на олімпіаду з математики студента можна вибрати $n = 5 + 6 = 11$ способами.

Відповідь: із двох груп на олімпіаду з математики студента можна вибрати 11-ма способами.

Правило множення. Якщо об'єкт A_1 можна вибрати n_1 способами, об'єкт A_2 – іншими n_2 способами, A_3 – відмінними від перших двох n_3 способами і т.д., A_k – n_k способами, то вибір упорядкованої системи об'єктів A_1, A_2, … і A_k можна здійснити $n_1 \cdot n_2 \cdot ... \cdot n_k$ способами.

Приклад 1.6. Студентка збирається на побачення і хоче вибрати вбрання, маючи три сукні, чотири намиста і дві пари взуття. Скількома різними способами вона це може зробити?

Розв’язання. Сукню студентка може обрати 3-ма способами, намисто – 4-ма способами, взуття – 2-ма способами. Отже, кількість можливих видів її вбрання з сукні, намиста і взуття є $n = 3 \cdot 2 \cdot 4 = 24$ способів.

Відповідь: студентка має 24 способи різних видів вбрання із сукні, намиста і взуття.

1.4.2. Формули комбінаторики

Означення. Підмножини, складені з будь-яких елементів даної множини, які відрізняються елементами або порядком цих елементів, називаються *сполуками*.

Означення. Розміщеннями із n елементів по m називаються такі сполуки, які складаються з m елементів, взятих із даних n елементів $(m \le n)$ і відрізняються або порядком, або принаймні одним елементом.

Число розміщень із n елементів по m позначається через A_n^m і обчислюється за формулою

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

де $n! = 1 \cdot 2 \cdot 3 \cdot \ldots \cdot n$ – добуток n перших натуральних чисел.

Для обчислення розміщень використовуються:

- статистична функція **ПЕРЕСТ(*число*;*число_выбранных*)**, де *число* – кількість елементів, *число_выбранных* – кількість елементів у перестановці. Також розміщення можна обчислювати безпосередньо за формулою, скориставшись математичною функцією **ФАКТ(*число*)**, де *число* – невід’ємне число, для якого обчислюється факторіал;

- функція **permut(*n*,*m*)**, де *n* – кількість елементів, *m* – кількість елементів у перестановці; також розміщення можна обчислювати безпосередньо за формулою, скориставшись оператором для обчислення факторіала, який знаходиться на підпанелі **Calculator (Калькулятор)** панелі інструментів **Math (Математика)**.

<u>Приклад 1.7.</u> У групі 30 студентів. Скількома різними способами можна вибрати старосту, замісника старости і профорга, якщо один студент може займати тільки одну посаду?

<u>Розв'язання.</u> Потрібно знайти число сполук по 3 елемента, взятих із даних 30 елементів, для яких порядок вибору має значення, тобто число розміщень із 30 елементів по 3:

$$A_{30}^{3}=\frac{30!}{(30-3)!}=30\cdot 29\cdot 28=24360.$$

B3 f_x =ПЕРЕСТ(B1;B2)

	A	B	C	D
1	*n*	30		
2	*m*	3		
3		24360		

$$n := 30 \qquad m := 3$$

$$\mathrm{permut}(n,m) = 2.436\times 10^{4}$$

<u>Відповідь:</u> старосту, замісника старости і профорга можна вибрати 24360-ма способами.

Якщо $m=n$, то такі розміщення називаються *перестановками*, число таких розміщень позначається P_n і обчислюється за формулою $P_n=n!$.

Для обчислення перестановок використовуються:

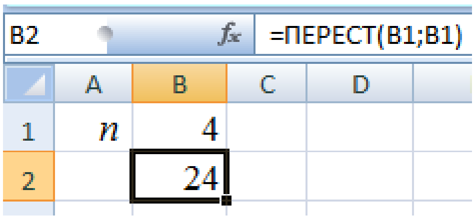
статистична функція **ПЕРЕСТ(*число*;*число_выбранных*)** або математична функція **ФАКТР(*число*)**, де *число* – кількість елементів, *число_выбранных* – кількість елементів у перестановці (у даному випадку *число*=*число_выбранных*);

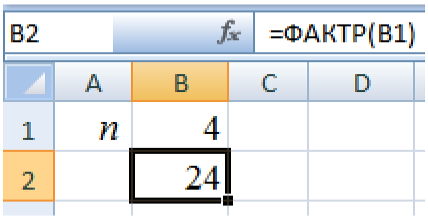
як зазначало раніше, на підпанелі **Calculator (Калькулятор)** панелі інструментів **Math (Математика)** є оператор для обчислення факторіала.

<u>**Приклад 1.8.**</u> Скількома різними способами можна розсадити за столом 4-х гостей на 4-х стільцях?

<u>Розв’язання.</u> Є 4 стільця і 4 гостя, отже, потрібно знайти число всіх перестановок із 4-х елементів:

$$P_4 = 4! = 4 \cdot 3 \cdot 2 \cdot 1 = 24.$$





$$n := 4$$
$$n! = 24$$

<u>Відповідь:</u> 4-х гостей на 4-х стільцях можна розсадити за столом 24-ма способами.

<u>*Означення.*</u> *Сполученнями (комбінацією) із* n *елементів по* m називаються такі сполуки, які складаються з m елементів, взятих із даних n елементів $(m \leq n)$ і відрізняються принаймні одним елементом.

Число сполучень із n елементів по m позначається через C_n^m і обчислюється за формулою

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

Для обчислення сполучень використовуються:

математична функція **ЧИСЛКОМБ(*число*;*число_выбранных*)**, де *число* – кількість елементів, *число_выбранных* – кількість елементів у комбінації;

функція **combin(*n*,*m*)**, де n – кількість елементів, m – кількість елементів у комбінації.

Приклад 1.9. У турнірі беруть участь 8 шахістів, причому кожні два з них зустрічаються один раз. Скільки планується провести партій?

Розв'язання. Потрібно знайти кількість сполук, що складаються з 2-х елементів із даних 8-ми, які відрізняються принаймні одним елементом, тому

$$C_8^2 = \frac{8!}{(8-2)!2!} = \frac{7 \cdot 8}{2} = 28.$$

B3 fx =ЧИСЛКОМБ(B1;B2)

	A	B	C	D	E
1	*n*	8			
2	*m*	2			
3		28			

$n := 8 \quad m := 2$

$\text{combin}(n, m) = 28$

Відповідь: планується провести 28 партій.

Практичне заняття 1

1.1. Бухгалтери трьох фірм здають звіти у податкову інспекцію. Записати події, які означають, що звіт здано: 1) лише бухгалтером I фірми; 2) лише бухгалтерами II і III фірм; 3) бухгалтерами усіх трьох фірм; 4) бухгалтером принаймні однієї фірми; 5) бухгалтерами принаймні двох фірм; 6) бухгалтером лише однієї фірми; 7) бухгалтерами лише двох фірм; 8) бухгалтерами не більше двох фірм; 9) жоден бухгалтер не здав.

1.2. На підприємстві працює комплекс із трьох автоматизованих технологічних ліній. На пункт контролю передається сигнал про безвідмовну роботу кожної з ліній або про відмову якоїсь із них у роботі. Описати події, що передано сигнал: 1) про безвідмовну усіх ліній комплексу; 2) про відмову I і III ліній; 3) про відмову комплексу.

1.3. Для контролю відвідування студентами лекцій, деканат відібрав 5 особових справ. Кожний із 5-ти відібраних студентів може бути як присутнім, так і відсутнім на лекції. Розглянемо події: A_1 – лекцію відвідав хоча б один студент із відібраних; A_2 – лекцію відвідав тільки один студент із відібраних; A_3 – лекцію відвідали не менше 3 студентів із відібраних; A_4 – лекцію відвідали тільки 3 студента із відібраних; A_5 – лекцію відвідали усі відібрані студенти. Описати події: 1) $A_1 + A_2$; 2) $A_1 A_2$; 3) $A_2 + A_3$; 4) $A_3 + A_4 + A_5$; 5) $A_4 A_5$.

1.4. На лекції присутні 20 студентів однієї групи і 15 – іншої, потрібно вибрати 2-х чергових (по одному з кожної групи). Скільки існує способів їх вибрати?

1.5. У ліфт 16-ти поверхового будинку на першому поверсі зайшли 3 чоловік. Скількома способами вони можуть розподілитись по поверхам будинку, починаючи з другого?

1.6. У студентській їдальні у меню є 4 салати, 3 перших страви, 5 – других та 7 видів напоїв. Скількома способами можна скласти обіднє меню із 4 страв?

1.7. Із корпусу А у корпус Б можна пройти 3-ма шляхами, із корпусу Б у корпус В – 2-ма. Скількома шляхами можна пройти із корпусу А у корпус В через корпус Б?

1.8. На секції наукової конференції виступатимуть 7 доповідачів. Скількома способами їх можна розмістити у списку?

1.9. Скількома способами можна розмістити 11 книг на полиці, якщо з них – один трьохтомник, примірники якого повинні стояти поряд?

1.10. Скількома способами можна розмістити 5 чоловік у легковому автомобілі (5 місць), якщо тільки один з них – водій?

1.11. Скількома способами можна скласти список із 18 студентів групи, так щоб староста і його замісник знаходились у списку поряд?

1.12. У групі 17 студентів. Скількома способами можна викликати до дошки 5-х з них протягом заняття?

1.13. Скільки можна скласти трьохзначних номерів квартир, у яких усі цифри різні?

1.14. На кафедрі працює 15 викладачів. Скількома способами із них може бути складена екзаменаційна комісія, якщо до її складу повинно увійти 4 викладачі?

1.15. У технологічному відділі підприємства працює 5 технологів. Для підвищення кваліфікації 3-х з них є можливість відправити на курси. Скількома способами це можна зробити?

1.16. На рекламний відділ підприємства, у якому працює 9 співробітників, виділено 4 путівки. Скількома способами їх можна розподілити, якщо вони: а) різні; б) однакові.

1.17. На контроль із кондитерського цеху поступило 35 тортів, 5 з них не відповідають стандарту. Контролер бере на перевірку 7 тортів. Скільки існує способів таких, що із 7-ми взятих тортів а) жодного бракованого; б) 3 бракованих; в) 5 бракованих.

1.18. Для проведення рольової гри групу з 19 студентів потрібно розділити на три групи по 5, 10 і 4. Скількома способами це можна зробити?

1.19. У корзині лежать 20 яблук і 15 груш. Яка ймовірність дістати з корзини: 1) одне яблуко; 2) одну грушу; 3) яблуко або грушу; 4) одну сливу.

1.20. У ящику 11 куль із номерами від 1 до 11. Дістали 1 кулю. Яка ймовірність, що її номер 1) 7; 2) не перевищує 5; 3) не перевищує 11; 4) 12.

1.21. Монету підкидають двічі. Знайти ймовірність того, що хоча б один раз випаде герб.

1.22. Із 20 акціонерних товариств 4 є банкротами. Клієнт придбав по одній акції 6 акціонерних товариств. Яка ймовірність того, що серед куплених акцій 2 виявляться акціями банкротів?

1.23. У коробці 4 жовтих і 5 червоних маркерів. Випало 4 маркери. Яка ймовірність того, що 3 з них червоні?

1.24. В ящику 10 білих і 6 чорних ґудзиків. Навмання взято 2 ґудзики. Яка ймовірність того, що вони: а) білі; б) чорні; в) один білий і один чорний.

1.25. У групі 24 студенти, з них 14 дівчат. Для участі у науковій конференції було відібрано 4-х студентів. Яка ймовірність того, що серед відібраних 2 дівчини?

1.26. У маркетинговому відділі концерну працює 25 співробітників, 6 з них – чоловіки. Яка ймовірність того, що серед 5-ти навмання відібраних для проходження переатестації співробітників хоча б один чоловік?

1.27. У ящику 50 пиріжків, 5 з них без начинки. Продавець навмання взяв 3 пиріжки. Яка ймовірність того, що хоча б один з них без начинки?

1.28. Десять чоловік незалежно один від одного сідають у поїзд метро, який містить 10 вагонів. Знайти ймовірність того, що всі вони опиняться у різних вагонах?

1.29. У ліфт дев'ятиповерхового будинку на першому поверсі ввійшли чотири людини. Кожний із них рівно ймовірно, незалежно від інших, може вийти із ліфта на будь-якому поверсі, починаючи із другого. Знайти ймовірність того, що всі пасажири вийдуть із ліфта на: а) різних поверхах; б) одному поверсі.

1.30. Гардеробниця видала номерки чотирьом джентльменам, що здали свої циліндри, але потім переплутала головні убори і повісила їх навмання. Знайти ймовірність того, що: а) кожний джентльмен отримає свій циліндр; б) рівно три джентльмена отримають свій циліндр; в) тільки два джентльмени отримають свій головний убір; г) тільки один джентльмен отримає свій циліндр; д) ніхто не отримає свого циліндра.

1.31. Практика показала, що із 100 прогнозів, які робить експерт-аналітик рекламної фірми «Небо», 85 справджуються. Знайти відносну частоту правдивих прогнозів цього експерта-аналітика.

1.32. Стрілець зробив 10 пострілів, із них 8 – влучних. Знайти відносну частоту влучень стрільця.

1.33. Відомо, що у деякому регіоні відносна частота народження дівчаток дорівнює 0,49. Скільки діток народилось протягом минулого року, якщо дівчаток народилось 980?

1.34. При стрільбі із пістолета відносна частота влучень спортсмена становить 0,95. Зроблено 160 пострілів. Знайти число влучень.

1.35. Телефонна лінія з’єднує два пункти, що знаходяться на відстані 3 км. У невідомому місці вона пошкодилась. Знайти ймовірність того, що вона пошкодилась не далі, ніж 500 м від першого пункту.

1.36. Два постачальника мають привезти товар у один супермаркет. Час приїзду кожного з них незалежний і рівно можливий протягом доби. Визначити ймовірність того, що одному із постачальників прийдеться чекати, якщо час прийому товару від першого постачальника 40 хв., від другого – 1 год.

1.37. Із відрізка $[0;1]$ випадково вибрано два дійсних числа. Яка ймовірність того, що їх сума не більша за 1, а добуток не більший $2/9$?

1.38. Монета, товщиною 2 мм, має діаметр 20 мм. Яка ймовірність того, що при падінні вона стане на ребро?

1.39. На клавіатуру комп’ютера капнула і розтеклася крапля кави, утворивши круг радіуса r см. Яка ймовірність того, що вона не протече між кнопками, якщо вони мають форму квадрата зі стороною a см.

Лекція 2. Теореми додавання та множення ймовірностей

2.1. Теорема додавання ймовірностей

2.2. Залежні та незалежні випадкові події. Теорема множення ймовірностей

2.3. Формула повної ймовірності та формули Байєса

2.1. Теорема додавання ймовірностей

Теорема 2.1. Ймовірність суми двох несумісних подій *A* і *B* дорівнює сумі ймовірностей цих подій:

$$P(A+B)=P(A)+P(B) \tag{2.1}$$

Доведення. Нехай n – число всіх можливих елементарних наслідків випробування, m_1 – число наслідків, що сприяють появі події *A*, m_2 – число наслідків, що сприяють появі події *B*. Із того, що події *A* і *B* несумісні випливає, що жоден наслідок, при якому відбудеться подія *A* не буде сприяти появі події *B*, тому події $A+B$ сприятимуть m_1+m_2 наслідків. За класичним означенням ймовірності

$$P(A+B)=\frac{m_1+m_2}{n}=\frac{m_1}{n}+\frac{m_2}{n}=P(A)+P(B). \bullet$$

Наслідок 2.1. Якщо випадкові події $A_1, A_2, ..., A_n$ попарно несумісні, то ймовірність суми цих подій дорівнює сумі їх ймовірносте й

$$P(A_1+A_2+...+A_n)=P(A_1)+P(A_2)+...+P(A_n). \tag{2.2}$$

Наслідок 2.2. Сума ймовірностей попарно несумісних подій $A_1, A_2, ..., A_n$, що утворюють повну групу подій, дорівнює 1

$$P(A_1)+P(A_2)+...+P(A_n)=1. \tag{2.3}$$

Наслідок 2.3. Сума ймовірностей протилежних подій A і $\bar{A}$ дорівнює 1

$$P(A)+P(\bar{A})=1. \tag{2.4}$$

<u>Приклад 2.1.</u> У лотереї розігрується 200 білетів, при цьому на 3 білети припадають виграші по 500 грн., на 10 білетів – по 20 грн., на 100 білетів – по 10 грн. Усі інші білети невиграшні. Студент купив 1 білет. Яка ймовірність, що він виграє не менше 20 грн.?

<u>Розв’язання.</u> Нехай подія A полягає у тому, що виграш складе 500 грн., подія B – виграш складе 20 грн. Тоді подія $A+B$ означатиме, що студент виграє не менше 20 грн. Події A і B несумісні, тому згідно з формулою (2.1):

$$P(A+B)=P(A)+P(B)=\frac{3}{200}+\frac{10}{200}=\frac{13}{200}=0,065.$$

<u>Відповідь:</u> ймовірність того, що студент виграє не менше 20 грн. складає 0,065.

<u>Приклад 2.2.</u> В ящику 14 деталей, з них 9 стандартних і 5 нестандартних. Знайти ймовірність того, що серед 4-х навмання взятих деталей хоча б одна виявиться стандартною.

<u>Розв’язання.</u> Нехай подія A полягає в тому, що серед 4-х навмання взятих деталей хоча б одна стандартна, тоді протилежна подія $\overline{A}$ – всі 4 взяті деталі нестандартні. З формули (2.4)

$$P(A)=1-P(\overline{A}).$$

Оскільки $P(\overline{A})=\dfrac{C_5^4}{C_{14}^4}=\dfrac{5}{1007}\approx 0,005$, то $P(A)=1-0,005=0,995$.

D3 =1-ЧИСЛКОМБ(D1;D2)/ЧИСЛКОМБ(B1;B2)

	A	B	C	D	E	F	G
1	*n1*	14	*n2*	5			
2	*m1*	4	*m2*	4			
3				0,995			

У данному випадку також можна використати статистичну функцію **ГИПЕРГЕОМЕТ(*число_успехов_в_выборке*;*размер_выборки*;*число_успехов_в_совокупности*;*размер_совокупности*)**, де *число_успехов_в_выборке*, *размер_выборки* – аргументи, повязані із пошуком числа наслідків, що сприяють появі події; *число_успехов_в_совокупности, размер_совокупности* – аргументи, повязані з пошуком числа всіх можливих елементарних наслідків випробування.

Тоді

D3 f_x =1-ГИПЕРГЕОМЕТ(D2;D1;B2;B1)

	A	B	C	D	E
1	*n1*	14	*n2*	5	
2	*m1*	4	*m2*	4	
3				0,995	

$$n1 := 14 \quad m1 := 4$$

$$n2 := 5 \quad m2 := 4$$

$$1 - \frac{\mathrm{combin}(n2, m2)}{\mathrm{combin}(n1, m1)} = 0.995$$

<u>Відповідь:</u> ймовірність того, що серед 4-х навмання взятих деталей хоча б одна виявиться стандартною становитиме 0,995.

2.2. Залежні та незалежні випадкові події. Теорема множення ймовірностей

<u>Означення.</u> Дві події називаються *незалежними*, якщо ймовірність появи однієї з них не залежить від появи або непояви іншої.

Наприклад, нехай випробування полягає в тому, що двічі підкидають монету. Подія A_1 полягає у тому, що при першому підкиданні випав герб, подія A_2 – при другому підкиданні випав герб. Зрозуміло, що ймовірність появи події A_2 не залежить від того чи відбулася, чи не відбулася подія A_1. Отже, події A_1 і A_2 незалежні.

<u>Означення.</u> Дві події називаються *залежними*, якщо ймовірність появи однієї з них залежить від появи або непояви іншої.

<u>Означення.</u> Умовною ймовірністю $P(B/A)$ *або* $P_A(B)$ називається ймовірність появи події *B* за умови, що подія *A* відбулася.

<u>Приклад 2.3.</u> У сенаті США 100 сенаторів, з них 58 республіканців та 42 демократа. Проведене голосування по деякому питанню. Його результати подані у вигляді таблиці

	Республіканці	Демократи	Усього
За	32	20	52
Проти	26	22	48
Усього	58	42	100

Навмання відібрано одного сенатора.

I. Знайти ймовірність того, що відібраний сенатор 1) республіканець; 2) демократ; 3) проголосував «за»; 4) проголосував «проти».

II. Він проголосував «за». Яка ймовірність того, що це був 1) республіканець; 2) демократ.

III. Яка ймовірність, що сенатор проголосував «за», якщо відомо, що він 1) республіканець; 2) демократ.

<u>Розв'язання.</u> Нехай подія A полягає у тому, що відібраний сенатор є республіканцем, B – демократом, C – відібраний сенатор проголосував «за», D – проголосував «проти».

Використовуючи таблицю даних, можна дати відповідь на питання частини I:

$$1)\ P(A)=\frac{58}{100};\ 2)\ P(B)=\frac{42}{100};\ 3)\ P(C)=\frac{52}{100};\ 4)\ P(D)=\frac{48}{100}.$$

У частині II відомо, що відбулась подія C. Потрібно знайти умовні ймовірності $P_C(A)$ і $P_C(B)$. Із таблиці даних відомо, що усього проголосувало «за» 52 сенатора, з них 32 республіканця та 20 демократів, отже,

$$1)\ P_C(A)=\frac{32}{52};\ 2)\ P_C(B)=\frac{20}{52}.$$

У частині III потрібно знайти умовні ймовірності $P_A(C)$ і $P_B(C)$. Усього республіканців 58 республіканців, із них 32 проголосувало «за». Отже, $P_A(C)=\frac{32}{58}$. Аналогічно обчислюється $P_B(C)=\frac{20}{42}$.

<u>Відповідь:</u> I. 1) $\frac{58}{100}$; 2) $\frac{42}{100}$; 3) $\frac{52}{100}$; 4) $\frac{48}{100}$.

II. 1) $\frac{32}{52}$; 2) $\frac{20}{52}$.

III. 1) $\frac{32}{58}$; 2) $\frac{20}{42}$.

Зауваження. Якщо події *A* і *B* незалежні, то умовна ймовірність дорівнює безумовній, тобто $P_A(B)=P(B)$.

<u>Теорема 2.2.</u> Ймовірність добутку двох випадкових подій *A* і *B* дорівнює добутку ймовірності однієї з них на умовну ймовірність іншої за умови, що перша подія відбулася

$$P(AB)=P(A)\cdot P_A(B)=P(B)\cdot P_B(A). \qquad (2.5)$$

<u>Доведення.</u> Нехай із загального числа n наслідків випробування події *A* сприяють m наслідків, події B – k наслідків, а сумісній появі подій AB – l наслідків ($l\le m$; $l\le k$) (рис. 2.1).

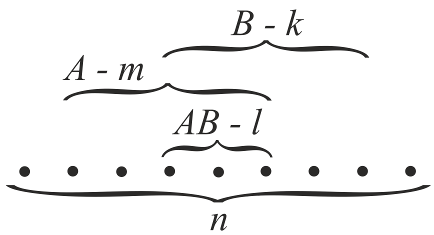


Рис. 2.1

Тоді згідно класичного означення ймовірності

$$P(A)=\frac{m}{n};\ P(AB)=\frac{l}{n}.$$

З рис. 2.1 видно: якщо подія А відбулася, то події В сприяють l наслідків, тому $P_A(B)=\dfrac{l}{m}$.

Отже, $P(A)\cdot P_A(B)=\dfrac{m}{n}\cdot\dfrac{l}{m}=\dfrac{l}{n}$, що і доводить першу частину формули (2.5). Аналогічно доводиться і друга частина цієї формули.•

<u>Наслідок 2.4.</u> Якщо A і B дві незалежні події, то ймовірність сумісної їх появи дорівнює добутку ймовірностей цих подій

$$P(AB)=P(A)\cdot P(B).$$

Зауваження. Теорема 2.2 справджується і для n довільних подій $A_1, A_2, ..., A_n$, тобто

$$P(A_1A_2...A_n)=P(A_1)\cdot P_{A_1}(A_2)\cdot ...\cdot P_{A_1A_2...A_{n-1}}(A_n).$$

<u>Приклад 2.4.</u> У прикладі 2.3 знайти ймовірність того, що випадково відібраний сенатор виявиться республіканцем, який проголосував «за».

<u>Розв'язання.</u> Потрібно знайти ймовірність сумісної появи подій A і C (див. розв'язання приклад 2.3). За формулою (2.5):

$$P(AC)=P(A)\cdot P_A(C),$$

де $P(A)=\dfrac{58}{100}$; $P_A(C)=\dfrac{32}{58}$.

Отже, $P(AC)=\dfrac{58}{100}\cdot\dfrac{32}{58}=0,32$.

<u>Відповідь:</u> ймовірність того, що випадково відібраний сенатор виявиться республіканцем, який проголосував «за» становить 0,32.

<u>Приклад 2.5.</u> Три стрільця стріляють по мішені. Ймовірність, що влучить у мішень перший – 0,8; другий – 0,9; третій – 0,7. Зроблено один залп, тобто усі три стрільця зробили одночасно по одному пострілу. Яка ймовірність, що у мішені буде хоча б одна пробоїна?

Розв’язання. Нехай подія A_i $(i=\overline{1,3})$ полягає у тому, що $i-$тий стрілець влучив у мішень, $\overline{A}_i$ $(i=\overline{1,3})$ – $i-$тий стрілець не влучив у мішень, A – у мішені буде хоча б одна пробоїна. Протилежною до події A буде подія $\overline{A}$ – у мішені немає жодної пробоїни, яка є сумісною з появою подій $\overline{A}_1$, $\overline{A}_2$, $\overline{A}_3$, тобто $\overline{A}=\overline{A}_1\overline{A}_2\overline{A}_3$. Згідно з формулою (2.4)

$$P(A)=1-P(\overline{A})=1-P(\overline{A}_1\overline{A}_2\overline{A}_3).$$

Події $\overline{A}_1$, $\overline{A}_2$, $\overline{A}_3$ – незалежні, тому

$$P(\overline{A}_1\overline{A}_2\overline{A}_3)=P(\overline{A}_1)P(\overline{A}_2)P(\overline{A}_3)=(1-P(A_1))(1-P(A_2))(1-P(A_3))=$$
$$=(1-0,8)(1-0,9)(1-0,7)=0,2\cdot 0,1\cdot 0,3=0,006$$

Отже, шукана ймовірність

$$P(A)=1-0,006=0,994.$$

Відповідь: ймовірність, що у мішені буде хоча б одна пробоїна дорівнює 0,994.

Теорема 2.3. Якщо випадкові події А і В сумісні, то

$$P(A+B)=P(A)+P(B)-P(AB). \qquad (2.6)$$

Доведення. Подія $A+B$ відбудеться, якщо відбудеться одна із несумісних подій $A\overline{B}$, $\overline{A}B$, AB. Отже, подію $A+B$ можна записати як суму цих трьох несумісних подій:

$$A+B=A\overline{B}+\overline{A}B+AB.$$

За теоремою додавання ймовірностей несумісних подій

$$P(A+B)=P(A\overline{B})+P(\overline{A}B)+P(AB). \qquad (2.7)$$

Враховуючи, що $A=A\overline{B}+AB$, $P(A)=P(A\overline{B})+P(AB)$, звідки $P(A\overline{B})=P(A)-P(AB)$ і аналогічно $P(\overline{A}B)=P(B)-P(AB)$. Підставивши знайдені вирази у формулу (2.7) можна отримати рівність (2.6).•

2.3. Формула повної ймовірності та формули Байєса

Нехай події H_1, H_2, …, H_n, що називаються *гіпотезами*, утворюють повну групу попарно несумісних подій.

<u>Теорема 2.4.</u> Якщо подія *A* може відбутися лише за умов появи однієї з подій H_1, H_2, …, H_n, які утворюють повну групу попарно несумісних подій, то ймовірність події *A* дорівнює

$$P(A)=\sum_{i=1}^{n}P(H_i)P_{H_i}(A), \tag{2.8}$$

де $P(H_i)$ – ймовірність гіпотези H_i, $i=\overline{1,n}$; $P_{H_i}(A)$ – умовна ймовірність події *A* за умови, що подія H_i відбулася.

<u>Доведення.</u> Згідно з умовою теореми подія А настане, якщо настане одна з несумісних подій $AH_1,\ AH_2,\ \ldots,\ AH_n,$ тобто $A=H_1A+H_2A+\ldots+H_nA.$

За теоремою додавання ймовірностей несумісних подій

$$P(A)=P(H_1A)+P(H_2A)+\ldots+P(H_nA)=\sum_{i=1}^{n}P(H_iA).$$

З теореми множення ймовірностей $P(H_iA)=P(H_i)P_{H_i}(A).$ Враховуючи цю рівність

$$P(A)=\sum_{i=1}^{n}P(H_i)P_{H_i}(A).\ \bullet$$

Формула (2.8) називається *формулою повної ймовірності*.

<u>Приклад 2.6.</u> У торгівельну фірму поступили пральні машини від трьох постачальників у відношенні $1:4:5$. Практичний досвід показав, що пральні машини, які поступили від І-го, ІІ-го і ІІІ-го постачальників, не потребують ремонту протягом гарантійного терміна відповідно у 98 %, 88 % і 92 % випадках. Знайти ймовірність того, що пральна машина, придбана у торгівельній фірмі, не потребуватиме ремонту протягом гарантійного терміну.

<u>Розв'язання.</u> Нехай події H_i $(i=\overline{1,3})$ полягають у тому, що пральна машина поступила у торгівельну фірму від $i-$го постачальника. *A* – пральна

машина, придбана у торгівельній фірмі, не потребуватиме ремонту протягом гарантійного терміну.

За умовою задачі

$$P(H_1)=\frac{1}{10};\ P_{H_1}(A)=0,98;$$

$$P(H_2)=\frac{4}{10};\ P_{H_2}(A)=0,88;$$

$$P(H_3)=\frac{5}{10};\ P_{H_3}(A)=0,92.$$

Згідно з формулою повної ймовірності

$$P(A)=P(H_1)P_{H_1}(A)+P(H_2)P_{H_2}(A)+P(H_3)P_{H_3}(A)=0,1\cdot 0,98+0,4\cdot 0,88+$$

$$+0,5\cdot 0,92=0,91.$$

Відповідь: ймовірність того, що пральна машина, придбана у торгівельній фірмі, не потребуватиме ремонту протягом гарантійного терміну 0,91.

Наслідком теореми множення ймовірностей і формули повної ймовірності є формула Байєса. Вона застосовується, коли подія *A* може відбутися тільки з однією із гіпотез $H_1,\ H_2,\ \ldots,\ H_n$, які утворюють повну групу подій, і відомо, що подія *A* відбулася. Необхідно провести кількісну переоцінку апріорних (відомих до випробувань) ймовірностей цих гіпотез $P(H_1),\ P(H_2),\ \ldots,\ P(H_n)$, тобто потрібно знайти апостеріорні (отримані після проведення випробування) умовні ймовірності гіпотез $P_A(H_1),\ P_A(H_2),\ \ldots,\ P_A(H_n)$.

Для знаходження такої формули потрібно записати теорему множення ймовірностей

$$P(AH_i)=P(A)P_{H_i}(A).$$

Звідси

$$P_A(H_i)=\frac{P(A)P_{H_i}(A)}{\sum_{i=1}^{n}P(H_i)P_{H_i}(A)}. \tag{2.9}$$

Формули (2.9) називаються *формулами Байєса*.

<u>Приклад 2.7.</u> Виконується умова приклада 2.6. Торгівельна фірма продала пральну машину, яка потребувала ремонту протягом гарантійного терміну. Яка ймовірність, що ця пральна машина поступила від І-го постачальника?

<u>Розв'язання.</u> Із введених у прикладі 2.6 позначень, подія $\overline{A}$ полягатиме у тому, що пральна машина потребує ремонту у гарантійний термін. Тоді

$$P(\overline{A}) = 1 - P(A) = 1 - 0{,}91 = 0{,}09 \text{ (див. приклад 2.6).}$$

З умови прикладу 2.6:

$$P_{H_1}(\overline{A}) = 1 - 0{,}98 = 0{,}02;$$

$$P_{H_2}(\overline{A}) = 1 - 0{,}88 = 0{,}12;$$

$$P_{H_3}(\overline{A}) = 1 - 0{,}92 = 0{,}08.$$

За формулою Байєса

$$P_{\overline{A}}(H_1) = \frac{0{,}1 \cdot 0{,}02}{0{,}09} \approx 0{,}022.$$

<u>Відповідь:</u> ймовірність, що пральна машина, яка потребувала ремонту протягом гарантійного терміну, поступила від І-го постачальника становить 0,022.

Практичне заняття 2

2.1. В отриманій підприємством партії товарів міститься 200 виробів І-го сорту, 100 – ІІ-го сорту, 30 – ІІІ-го сорту. Знайти ймовірність того, що навмання взятий виріб виявиться а) ІІ-го або ІІІ-го сорту, б) І-го сорту.

2.2. На свято в рекламну компанію замовили 20 коробок цукерок (11 – «Прометей», решта – «Вечірній Київ»). Визначити ймовірність того, що із 3-х навмання взятих коробок цукерок усі виявляться одного сорту.

2.3. Чотирьох працівників підприємства відправили на проходження переатестації. Ймовірність того, що І-й працівник переатестується успішно 0,6;

II – 0,8; III – 0,7; IV – 0,9. Знайти ймовірність того, що всі працівники підприємства переатестуються успішно.

2.4. В обмінному пункті є 10 банкнот номіналом 100 $, 4 з них нові. Працівник обмінного пункту навмання бере одна за одною 2 купюри. Яка ймовірність того, що обидві купюри нові?

2.5. У відділі кадрів знаходяться особові справи 33 працівників підприємства, 5 з них заповнені неповністю. Перевіряючий бере одна за одною 3 справи. Яка ймовірність того, що а) усі вони заповнені повністю; б) тільки третя за порядком – заповнена неповністю.

2.6. На підприємстві комплектують економічний відділ. Є 3 претендента. Ймовірність того, що приймуть І-го – 0,5; ІІ-го – 0,8; ІІІ-го – 0,9. Знайти ймовірність того, що на роботу приймуть 1) тільки ІІ-го претендента; 2) тільки одного претендента; 3) усіх трьох претендентів; 4) хоча б одного претендента.

2.7. В одному кошику лежать 5 яблук і 15 груш, у другому – 10 яблук і 12 слив. Із кожної корзини навмання взято по одному фрукту. Знайти ймовірність того, що взято а) грушу і сливу; б) два яблука.

2.8. У першій коробці лежать 7 білих і 11 червоних кульок, у другій – 11 білих і 7 червоних. Із кожної коробки навмання взято по одній кульці. Знайти ймовірність того, що хоча б одна з них біла.

2.9. Двом студентам доручили купити два торти на свято. Вони купували їх у різних магазинах, не домовляючись. Перший вибирав із 3 тортів «Вишня» і 5 – «Абрикос», другий – із 4 «Апельсин» і 5 «Вишня». Яка ймовірність того, що вони куплять однакові торти?

2.10. Ймовірність несправності комп'ютера керівника маркетингового відділу протягом дня становить 0,01. Яка ймовірність того, що протягом 5-ти днів у керівника відділу не виникнуть проблеми з комп'ютером?

2.11. Покупець шукає потрібний йому товар, відвідуючи для цього 4 магазини. Ймовірність того, що потрібний товар є в кожному із магазинів становить 0,3. Що ймовірніше: знайти шукану річ чи ні?

2.12. У лабораторію для аналізу поступило 15 пляшок солодкої води. Відомо, що 5 з них «Буратіно», 7 – «Лимонад», 3 – «Сітро». Експерт навмання бере 3 пляшки. Яка ймовірність того, що всі вони різні?

2.13. Охоронець має 7 ключів. Тільки один ключ підходить до замка, який він хоче відкрити, тому і перебирає ключі по черзі один за одним. Яка ймовірність того, що охоронець відчинить замок третім ключем?

2.14. Автоматизований комплекс по розфасовці рафінованого цукру складається із 3 незалежно працюючих ліній. Ймовірність несправності кожної з них відповідно 0,1; 0,02 та 0,03. Знайти ймовірність того, що а) усі 3 лінії працюють справно, б) хоча б одна з ліній вийде з ладу.

2.15. Відомо, що паспорт знаходиться у письмовому столі з ймовірністю 0,9; причому з однаковою ймовірністю він може міститись у одному із шести ящиків. Було перевірено п'ять ящиків, але паспорт не знайдено. Яка ймовірність того, що він лежить у шостому ящику?

2.16. Дві студентки Марина і Галина проживають в одній кімнаті у гуртожитку. Їх пригостили 2 шоколадними цукерками і 4 карамельками зі схожими обгортками. Дівчата по черзі беруть і з'їдають по одній цукерці (починає Марина). Вони домовились, що та, якій першій дістанеться шоколадна цукерка, буде прибирати у кімнаті. Яка ймовірність того, що кімнату доведеться прибирати Галині?

2.17. На підприємство поступила сировина від 3-х постачальників. Від I-го – 20 %, від II-го – 30 %, від III-го – 50 %. Відомо, що у I-го постачальника 90 % сировини вищого ґатунку, у II-го – 95 %, у III-го – 80 %. Знайти ймовірність того, що навмання використана сировина виявиться вищого ґатунку.

2.18. Три автомати розливають сік у банки. Перший автомат видає 2 % браку, другий – 7 %, третій – 10 %. Продуктивність першого автомату в 3 рази вища за продуктивність другого, а третього – у 2 рази менша, ніж другого. Навмання беруть банку із соком. Яка ймовірність, що вона бракована?

2.19. Співробітники відділу маркетинга припускають, що найближчим часом можна чекати зростання попиту на продукцію підприємства. Ймовірність цього вони оцінюють у 80 %. Консультаційна фірма, яка займається прогнозом стану ринку, підтвердила дане припущення. Практика показує, що позитивні прогнози консультаційної фірми збуваються із ймовірність 95 %, а негативні – з ймовірністю 99 %. Яка ймовірність того, що зростання попиту відбудеться?

2.20. Із 1000 ламп, що надійшли для продажу у супермаркет, 380 належать першій партії, 270 – другій, решта – третій. Відомо, що перша партія містила 4 % браку, друга – 3 %, третя – 6 %. Консультант взяв одну лампу. Визначити ймовірність того, що вона не бракована.

2.21. Співвідношення м’ясних страв до рибних, які пропонують ресторані, становить 3:1. Ймовірність отруєння м’ясною стравою 0,005, а рибною – 0,003. Яка ймовірність отруїтись, поївши у цьому ресторані?

2.22. У футбольній команді «Веселі равлики» 3 нападаючих. Ймовірність того, що перший попаде у ворота 0,4; другий – 0,5; третій – 0,6. При цьому перший б’є по воротам у 1,5 рази частіше, ніж другий, і у 2 рази частіше, ніж третій. Визначити, з якою ймовірністю «Веселі равлики» забивають гол.

2.23. У бібліотеці 60 % читачів – студенти економічних спеціальностей, решта – прикладних. Студенти-економісти беруть книги з економіки у 7 разів частіше, ніж з прикладних дисциплін, студенти-прикладники беруть книги з прикладних дисциплін у 9 разів частіше, ніж з економіки. Знайти ймовірність того, що видана у бібліотеці книга буде з економіки.

2.24. На відкритті студентської конференції були присутні викладачі і студенти університету у співвідношення 2:5. Ймовірність помилки при реєстрації викладача становила 0,05; студента – 0,15. Знайти ймовірність того, що навмання перевірена реєстраційна картка не містила помилки.

2.25. Студент може їхати в університет двома шляхами. Ймовірність попасти у «пробку» першим шляхом становить 0,2; другим – 0,1. Яка ймовірність того, що студент без перешкод добереться до університету?

2.26. Микола поклав 30 % коштів у банк «Приліт», 35 % – «Проліт», решту – склав у літрову банку. У кінці року кошти вкладені у «Приліт» збільшились на 15 %, у «Проліт» – на 8 %, у банці – не попсувались. На скільки відсотків збільшились доходи Миколи?

2.27. У магазині працюють два касири. Перший обслуговує у 1,5 рази більше покупців за зміну, ніж другий. Ймовірність помилки першого касира 0,04; другого – 0,03. Покупця було обраховано. Яка ймовірність того, що його обслуговував другий касир?

2.28. У банку працюють два автомати, які рахують купюри. Продуктивність першого у 2 рази вища, ніж другого. Ймовірність помилки першого – 0,01; другого – 0,005. При розрахунку клієнта банка була допущена помилка. Який автомат найімовірніше його розраховував?

2.29. Фермер доручив двом мисливцям застрілити вовка і пообіцяв премію за це у розмірі 200 $. Перший, більш досвідчений мисливець, попадає у звіря із ймовірністю 0,9; а другий – 0,6. Мисливці зустріли вовка і одночасно вистрілили. Вовк був убитий однією кулею. Як мисливцям розподілити премію?

2.30. Роман спіймав 7 карасів і 5 лящів. Поставив відро з рибою і поки відчиняв двері, кіт вкрав якусь одну рибину. Перша рибина, яку він витягнув чистити виявилась лящем. Яка ймовірність, що кіт вкрав карася?

2.31. При завантаженні комп’ютера може виникнути ряд проблем: 20 % з них пов’язані із програмним забезпеченням (в 1 із 50 випадків потрібно перевстановлювати операційну систему); 35 % – із вірусом (перевстановлення в 1 із 20 випадків); в інших випадках винуватці проблем – користувачі (перевстановлення в 1 із 30 випадків). Комп’ютер вийшов із ладу і вимагає перевстановлення. Яка із проблем найімовірніше це спричинила?

2.32. Іван для розв’язання складної задачі на іспиті вирішив скористатись дзвінком одному із чотирьох друзів. Слава і Максим можуть розв’язати задачу із ймовірністю 0,2; Петро – 0,5; Арсеній – 0,9. Перший, навмання вибраний друг, якому подзвонив Іван, правильно підказав розв’язання задачі. Яка ймовірність, що це був Арсеній?

2.33. Одного вихідного дня Василь вирішив запросити одну із своїх подружок у кіно. Юля погодиться з ймовірністю 0,1; Марина – 0,2; Світлана – 0,15. Василь випадковим чином набрав номер однієї із трьох своїх подружок і отримав відмову. Кому найімовірніше він зателефонував?

2.34. У будинку, у якому знаходиться офіс концерну «Понтік», розташовані два кафе «Ромашка» і «Пиріжок». 45 % співробітників концерну ходять обідати у кафе «Ромашка», решта – у кафе «Пиріжок». При цьому ймовірність того, що їх влаштує обіднє меню і вони пообідають становить відповідно 95 % і 90 %. Співробітнику концерну не вдалось пообідати. Яка ймовірність того, що він намагався пообідати у кафе «Ромашка»?

2.35. Один відомий володар, якому чаклун набрид своїми хибними передбаченнями, вирішив його стратити. Однак, як завжди поступають у такому випадку, вирішив дати йому останній шанс і запропонував розподілити по двом коробкам чотири кульки: дві білі і дві чорні. Палач навмання вибирає одну коробку і дістає звідти кульку. Якщо кулька виявиться білою чаклун залишається живим, інакше – його страчують. Яким чином чаклуну слід розмістити кульки в коробках, щоб збільшити ймовірність свого спасіння?

Лекція 3. Повторні незалежні випробування

3.1. Формула Бернуллі

3.2. Найвірогідніша частота

3.3. Граничні теореми схеми Бернуллі

3.1. Формула Бернуллі

Нехай проводиться n випробувань, ймовірність появи події A у кожному з них не залежить від результатів інших випробувань. У цьому випадку можна говорити, що розглядається послідовність незалежних випробувань відносно події A. Якщо ймовірність появи події A в послідовності незалежних випробувань набуває одного і того ж значення, то такі випробування називаються *схемою Бернуллі*.

Ймовірність того, що подія A відбудеться в кожному з n незалежних випробувань схеми Бернуллі позначається через $p = P(A)$, тоді ймовірність протилежної події $\overline{A}$, що подія A не відбудеться в кожному з незалежних випробувань позначається через q: $P(\overline{A}) = 1 - P(A) = 1 - p = q$.

Розглядається задача: обчислити ймовірність того, що подія A відбудеться рівно k разів (не має значення в якій послідовності) в n незалежних випробуваннях схеми Бернуллі.

Шукана ймовірність позначається через $P_n(k)$ і обчислюється за *формулою Бернуллі*

$$P_n(k) = C_n^k p^k q^{n-k}. \tag{3.1}$$

Зауваження. Формула (3.1) називається також формулою біноміального розподілу ймовірностей, тому що у її правій частині стоїть $(k+1)-$й член бінома Ньютона

$$(p+q)^n = C_n^0 q^n + C_n^1 p q^{n-1} + \ldots + C_n^k p^k q^{n-k} + \ldots + C_n^n p^n.$$

Доведення формули Бернуллі. Нехай подія B_k означає, що в n випробуваннях схеми Бернуллі подія A настане k разів, а, отже, не настане $n-k$ разів. Тоді подію B_k можна записати як суму несумісних подій:

$$B_k = \underbrace{AA...A}_{k}\underbrace{\overline{A}\overline{A}...\overline{A}}_{n-k} + ... + \underbrace{\overline{A}\overline{A}...\overline{A}}_{n-k}\underbrace{AA...A}_{k}. \qquad (3.2)$$

У виразі (3.2) кожний доданок має ймовірність $p^k q^{n-k}$ (згідно теореми множення ймовірностей). Кількість усіх доданків у цьому виразі дорівнює числу комбінацій з n елементів по k елементів, тобто $C_n^k = \dfrac{n!}{(n-k)!k!}$.

Кожен доданок у виразі (3.2) є несумісною подією з ймовірністю $p^k q^{n-k}$, отже,

$$P_n(k) = P(B_k) = C_n^k p^k q^{n-k}. \bullet$$

Для обчислення ймовірності за формулою Бернуллі використовуються:

статистична функція **БИНОМРАСП(*число_успехов*;*число_испытаний*;*вероятность_успеха*;*интегральная*)**, де *число_успехов* – кількість успішних наслідків у незалежних випробуваннях, *число_испытаний* – кількість незалежних випробувань, *вероятность_успеха* – ймовірність успіху у кожному з випробувань, *интегральная* – логічне значення, яке визначає вид функції, у данному випадку значення аргументу *интегральная* = «ЛОЖЬ» або «0» (загалом, у разі «ЛОЖЬ» або «0» – повертається значення функції ймовірнісної міри, тобто ймовірність того, що кількість успішних наслідків у незалежних випробуваннях дорівнює значенню аргументу *число_успехов*; у разі «ИСТИНА» або «1» – повертається значення інтегральної функції розподілу, тобто ймовірність того, що кількість успішних наслідків у незалежних випробуваннях не менша значення аргументу *число_успехов*);

функція **dbinom(*k*,*n*,*p*)**, де k – кількість успішних наслідків у незалежних випробуваннях, n – кількість незалежних випробувань, p – ймовірність успіху у кожному з випробувань.

Приклад 3.1. Ймовірність влучення в мішень при одному пострілі дорівнює 0,8. Знайти ймовірність 3-х влучень у мішень при 5-ти пострілах.

Розв'язання. Ймовірність влучення у мішень $p = 0{,}8$, тоді ймовірність промаху $q = 1 - p = 1 - 0{,}8 = 0{,}2$. Згідно формули Бернуллі:

$$P_5(3) = C_5^3 p^3 q^{5-3} = \frac{5!}{3!2!}(0{,}8)^3 (0{,}2)^2 = 0{,}2048.$$

B4 | =БИНОМРАСП(B2;B1;B3;ЛОЖЬ)

	A	B	C	D	E
1	*n*	5			
2	*k*	3			
3	*p*	0,8			
4		0,2048			

n := 5   k := 3   p := 0.8

dbinom(k, n, p) = 0.205

Відповідь: ймовірність 3-х влучень у мішень при 5-ти пострілах дорівнює 0,2048.

3.2. Найвірогідніша частота

Частота k_0 число настання події A в n незалежних випробуваннях називається *найвірогіднішою частотою* появи цієї події, якщо їй відповідає найбільша ймовірність.

Якщо k_0 найвірогідніша частота, то

$$P_n(k_0 - 1) \le P_n(k_0) \ge P_n(k_0 + 1),$$

тобто

$$\frac{P_n(k_0)}{P_n(k_0-1)} \geq 1, \ \frac{P_n(k_0)}{P_n(k_0+1)} \geq 1.$$

Застосувавши формулу Бернуллі, можна отримати

$$\frac{n-k_0+1}{k_0}\frac{p}{q} \geq 1, \ \frac{n-k_0}{k_0+1}\frac{p}{q} \leq 1.$$

Ці нерівності еквівалентні наступним нерівностям

$$np - k_0 p + p \geq k_0 q,$$

$$np - k_0 p \leq k_0 q + q$$

або

$$np + p \geq k_0 p + k_0 q = k_0(p+q) = k_0,$$

$$np - q \leq k_0 p + k_0 q = k_0(p+q) = k_0.$$

Звідки можна отримати остаточний діапазон зміни k_0 :

$$np - q \leq k_0 \leq np + p.$$

Сегмент $[np-q; np+p]$ має довжину рівну одиниці. Тому цей сегмент містить або одне ціле число, або (якщо кінці сегмента є цілі числа) два цілих числа, які і є найвірогіднішою частотою події *A*.

__Приклад 3.2.__ Гральний кубик підкидають 35 разів. Яка найвірогідніша частота появи грані з шести очками?

Розв’язання. За умовою $n=35, \ p=\frac{1}{6}, \ q=\frac{5}{6},$ тоді

$$np+p=\frac{35}{6}+\frac{1}{6}=6 \text{ і } np-q=\frac{35}{6}-\frac{5}{6}=5.$$

Отже, є дві найвірогідніші частоти $k_0=5$ і $k_0=6$.

Відповідь: найвірогідніші частоти появи грані з шести очками $k_0=5$ і $k_0=6$.

3.3. Граничні теореми схеми Бернуллі

Формули біноміального розподілу при великих n приводять до дуже громіздких обчислень. Тому важливо мати достатньо прості наближені формули для обчислення відповідних ймовірностей.

Для задач, в яких число незалежних випробувань велике, а ймовірність p появи події A в одному випробуванні мала, шукана ймовірність $P_n(k)$ може бути знайдена з достатньою точністю за формулою Пуассона.

Нехай n велике, p мала, а добуток $n \cdot p = \lambda$ є деяка фіксована величина. Враховуючи це, із формули Бернуллі можна отримати

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

$$= \left(\frac{n(n-1)...(n-k+1)}{n^k}\right) \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \left(1 - \frac{\lambda}{n}\right)^{-k}.$$

Так як при фіксованому k і при $n \to \infty$ $\lim\limits_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^n = e^{-\lambda}$, $\lim\limits_{n\to\infty}\left(1-\frac{\lambda}{n}\right)^{-k} = 1$ і

$$\lim_{n\to\infty} \frac{n(n-1)...(n-k+1)}{n^k} = \lim_{n\to\infty} \underbrace{\left(1-\frac{1}{n}\right)\left(1-\frac{2}{n}\right)...\left(1-\frac{k-1}{n}\right)}_{k-1} = 1, \text{ то}$$

$$\lim_{n\to\infty} P_n(k) = \frac{\lambda^k}{k!} e^{-\lambda}. \qquad (3.3)$$

Формула (3.3) називається *формулою Пуассона*.

Для практичного застосування при великих n і малих p вона застосовується як наближена рівність

$$P_n(k) \approx \frac{\lambda^k}{k!} e^{-\lambda}, \qquad (3.4)$$

де $\lambda = np.$

Для обчислення ймовірності за формулою Пуассона використовуються:

- статистична функція **ПУАССОН(*x*;*среднее*;*интегральная*)**, де *x* – кількість успішних наслідків у незалежних випробуваннях, *среднее* –

шукане чисельне значення, *интегральная* – логічне значення, що визначає тип кінцевого результату, у разі «ЛОЖЬ» або «0» – повертається значення щільності розподілу Пуассона, тобто ймовірність того, що кількість успішних наслідків у незалежних випробуваннях дорівнює значенню аргумента x; у разі «ИСТИНА» або «1» – повертається значення інтегральної функції розподілу Пуассона, тобто ймовірність того, що кількість успішних наслідків у незалежних випробуваннях не більше значення аргумента x, причому у даному випадку *среднее*= $\lambda = np$, *интегральная*=«ЛОЖЬ» або «0».

функція **dpois(*k*,λ)**, де k – кількість успішних випробувань, $\lambda = np$.

<u>Приклад 3.3.</u> Ймовірність виготовлення бракованого виробу 0,01. Яка ймовірність, що у партії зі 100 виробів, 2 – виявляться бракованими?

<u>Розв’язання.</u> За умовою $n = 100$, $p = 0{,}01$, $\lambda = np = 100 \cdot 0{,}01 = 1$, тоді згідно формули Пуассона

$$P_{100}(2) \approx \frac{1}{2} e^{-1} \approx 0{,}184.$$

B4 =ПУАССОН(B2;B1*B3;ЛОЖЬ)

	A	B	C	D	E
1	*n*	100			
2	*k*	2			
3	*p*	0,01			
4		0,18394			

$n := 100 \quad k := 2 \quad p := 0.01$

$\mathrm{dpois}(2, 100 \cdot 0.01) = 0.184$

<u>Відповідь:</u> ймовірність того, що у партії зі 100 виробів, 2 – виявляться бракованими становить 0,184.

Якщо кількість випробувань n велика, а ймовірність p не прямує до нуля, тоді число λ не можна вважати сталим і формула Пуассона не застосовується. У цьому випадку $P_n(k)$ обчислюється за наближеною формулою Муавра-Лапласа.

Теорема 3.1. (Локальна теорема Муавра-Лапласа). Якщо ймовірність p появи події А у кожному випробуванні стала величина і $0 < p < 1$, то ймовірність $P_n(k)$ того, що подія *А* з'явиться у n незалежних випробуваннях k разів наближено дорівнює

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \varphi(x), \tag{3.5}$$

де $$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \tag{3.6}$$

і $$x = \frac{k - np}{\sqrt{npq}}. \tag{3.7}$$

Функція $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ називається *функцією Гаусса*. Ця функція часто використовується у теорії ймовірностей, тому її значення наведені у відповідній таблиці (додаток 1). При використанні цієї таблиці необхідно мати на увазі властивості функції $\varphi(x)$.

Основні властивості функції Гаусса.

1. $\varphi(x)$ – парна, тобто $\varphi(-x) = \varphi(x)$.
2. $\max\limits_{x} \varphi(x) = \varphi(0) = \frac{1}{\sqrt{2\pi}}$.
3. $\lim\limits_{x \to \pm\infty} \varphi(x) = 0$.

Графік функції Гаусса зображено на рис. 3.1.

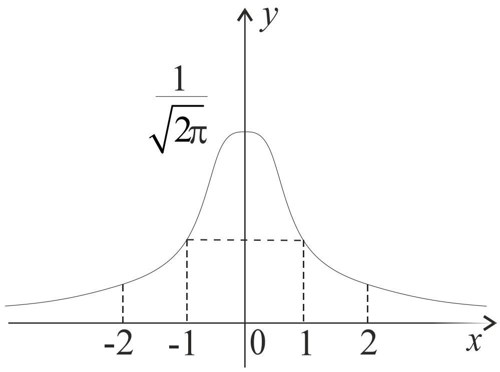


Рис. 3.1

Для обчислення ймовірності за допомогою локальної теореми Муавра-Лапласа використовуються:

- для обчислення $x = \frac{k - np}{\sqrt{npq}}$ – статистична функція **НОРМАЛИЗАЦИЯ(*x*;*среднее*;*стандартное_откл*)**, де *x* – кількість успішних наслідків у незалежних випробуваннях, *среднее* – середнє арифметичне значення розподілу, *стандартное_откл* – стандартне відхилення розподілу, причому *среднее*=np, *стандартное_ откл*= $\sqrt{npq}$;

- для знаходження значень функції Гаусса $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ – статистична функція **НОРМРАСП(*x*;*среднее*;*стандартное_откл*; *интегральная*)**, де *x* – значення, для якого будується розподіл, *среднее* – середнє арифметичне значення розподілу, *стандартное_откл* – стандартне відхилення розподілу; *интегральная* – логічне значення, яке визначає тип функції (у разі «ИСТИНА» або «1» повертається значення інтегральної функції розподілу, у разі «ЛОЖЬ» або «0» – значення щільності розподілу), причому *среднее*=0, *стандартное_откл*=1, *интегральная*=«ЛОЖЬ» або «0»;

для знаходження значень функції Гаусса – функція **dnorm(*x*, μ, σ)**, де x – значення, для якого будується розподіл, μ – середнє арифметичне значення розподілу, σ – стандартне відхилення розподілу, причому $\mu = 0, \ \sigma = 1$.

<u>Приклад 3.4.</u> Ймовірність народження хлопчика дорівнює 0,51. Знайти ймовірність того, що серед 100 немовлят виявиться 50 хлопчиків.

<u>Розв'язання.</u> За умовою $n = 100$, $p = 0,51$, $q = 1 - 0,51 = 0,49$, $k = 50$, тоді згідно з формулою Муавра-Лапласа

$$P_{100}(50) \approx \frac{1}{\sqrt{100 \cdot 0,51 \cdot 0,49}} \varphi(x),$$

де $x = \dfrac{50 - 100 \cdot 0,51}{\sqrt{100 \cdot 0,51 \cdot 0,49}} = \dfrac{-1}{5} = -0,2$.

Значення функції Гаусса $\varphi(-0,2)$ знаходиться з таблиці (додаток 1), враховуючи, що $\varphi(-0,2) = \varphi(0,2) = 0,391$.

Отже, $P_{100}(50) \approx \dfrac{0,391}{5} \approx 0,078$.

1) Розрахунки проведено поетапно:

B3 f_x =НОРМАЛИЗАЦИЯ(D1;B1*B2;КОРЕНЬ(B1*B2*D2))

	A	B	C	D	E	F	G
1	n	100	k	50			
2	p	0,51	q	0,49			
3	x	-0,2					

2)

B4 f_x =НОРМРАСП(B3;0;1;ЛОЖЬ)

	A	B	C	D	E
1	n	100	k	50	
2	p	0,51	q	0,49	
3	x	-0,2			
4	$\varphi(x)$	0,391			

3)

B5 f_x =(1/КОРЕНЬ(B1*B2*D2))*B4

	A	B	C	D	E
1	n	100	k	50	
2	p	0,51	q	0,49	
3	x	-0,2			
4	$\varphi(x)$	0,391			
5	$P_{100}(50)$	0,078			

$$n := 100 \qquad k := 50$$

$$p := 0.51 \qquad q := 1 - p = 0.49$$

$$x := \frac{k - n \cdot p}{\sqrt{n \cdot p \cdot q}} = -0.2$$

$$fi := dnorm(x, 0, 1) = 0.391$$

$$P := \frac{1}{\sqrt{n \cdot p \cdot q}} \cdot fi = 0.078$$

<u>Відповідь:</u> ймовірність того, що серед 100 немовлят виявиться 50 хлопчиків дорівнює 0,078.

Локальна теорема Муавра-Лапласа дозволяє знайти ймовірність того, що при великій кількості випробувань n подія A настане k разів.

У багатьох прикладних задачах більш суттєву роль грає оцінка ймовірності того, що число появ події А лежить у деякому діапазоні. Таку оцінку дає інтегральна теорема Муавра-Лапласа.

Нехай $P_n(k_1, k_2)$ ймовірність того, що в n незалежних випробуваннях схеми Бернуллі подія A настане не менше k_1 і не більше k_2 разів.

<u>Теорема 3.2. (Інтегральна теорема Муавра-Лапласа).</u> Якщо ймовірність p появи події A в кожному випробуванні і $0 < p < 1$, то ймовірність $P_n(k_1, k_2)$

того, що подія A настане не менше k_1 і не більше k_2 разів, при достатньо великій кількості випробувань n наближено дорівнює

$$P_n(k_1, k_2) \approx \Phi(x_2) - \Phi(x_1), \qquad (3.8)$$

де $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{z^2}{2}} dz$ – *функція Лапласа,*

$$x_1 = \frac{k_1 - np}{\sqrt{npq}}, \quad x_2 = \frac{k_2 - np}{\sqrt{npq}}.$$

Значення функції Лапласа $\Phi(x)$ наведені у таблиці (додаток 2). Для використання цієї таблиці потрібно знати властивості функції $\Phi(x)$.

Основні властивості функції Лапласа.

1. $\Phi(x)$ – непарна, тобто $\Phi(-x) = -\Phi(x)$.
2. $\Phi(0) = 0$.
3. $\Phi(x)$ зростає на $(-\infty; +\infty)$.
4. $\lim\limits_{x \to \infty} \Phi(x) = 0{,}5$.

Зауваження. Якщо $x > 5$, то $\Phi(x) = 0{,}5$.

Графік функції Лапласа зображено на рис. 3.2.

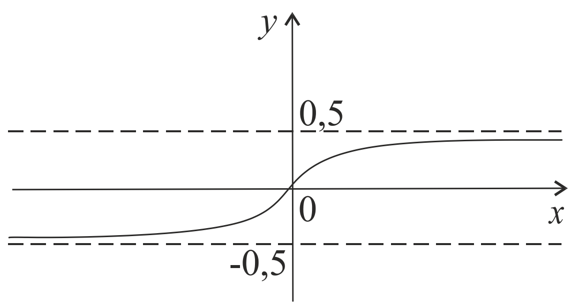


Рис. 3.2

Для обчислення ймовірності за допомогою інтегральної теореми Муавра-Лапласа використовуються:

для обчислення $x = \dfrac{k - np}{\sqrt{npq}}$ – статистична функція **НОРМАЛИЗАЦИЯ(*x*;*среднее*;*стандартное_откл*)**, опис приведено раніше; для знаходження значення функції Лапласа $\Phi(x) = \dfrac{1}{\sqrt{2\pi}} \int\limits_0^x e^{-\frac{z^2}{2}} dz$ – статистична функція **НОРМСТРАСП(*z*)-0,5**, де z – значення, для якого будується розподіл;

для знаходження значень функції Лапласа – функція **pnorm(*x*, μ ,σ) - 0,5**, де x – значення, для якого будується розподіл, μ – середнє арифметичне значення розподілу, σ – стандартне відхилення розподілу, причому $\mu = 0$, $\sigma = 1$; також можна використати функцію **cnorm(*x*)-0,5**, де x – значення, для якого будується розподіл.

<u>Приклад 3.5.</u> Ймовірність того, що студент в їдальні замовить першу страву дорівнює 0,8. Знайти ймовірність того, що в їдальні із 100 студентів першу страву замовлять: 1) не менше 75 і не більше 90 студентів; 2) не менше 75 студентів; 3) не більше 74 студентів.

<u>Розв’язання.</u> У цій задачі потрібно скористатись інтегральною теоремою Муавра-Лапласа. За умовою $n = 100$, $p = 0,8$, $q = 1 - 0,8 = 0,2$.

1) Оскільки $k_1 = 75$, $k_2 = 90$, то

$$x_1 = \frac{k_1 - np}{\sqrt{npq}} = \frac{75 - 100 \cdot 0,8}{\sqrt{100 \cdot 0,8 \cdot 0,2}} = -1,25, \; x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{90 - 100 \cdot 0,8}{\sqrt{100 \cdot 0,8 \cdot 0,2}} = 2,5.$$

Приймаючи до уваги непарність функції Лапласа:

$$P_{100}(75;90) = \Phi(2,5) - \Phi(-1,25) = \Phi(2,5) + \Phi(1,25).$$

З таблиці (додаток 2) $\Phi(2,5) = 0,4938$, $\Phi(1,25) = 0,3944$, тоді

$$P_{100}(75;90) = 0,4938 + 0,3944 = 0,8882.$$

2) Вимога, щоб першу страву замовили не менше 75 студентів із 100, означає, що $k_1 = 75$, $k_2 = 100$.

Тоді $x_1 = \dfrac{k_1 - np}{\sqrt{npq}} = \dfrac{75 - 100\cdot 0,8}{\sqrt{100\cdot 0,8\cdot 0,2}} = -1,25$, $x_2 = \dfrac{k_2 - np}{\sqrt{npq}} = \dfrac{100 - 100\cdot 0,8}{\sqrt{100\cdot 0,8\cdot 0,2}} = 5$.

Отже, шукана ймовірність

$$P_{100}(75;100) = \Phi(5) + \Phi(1,25) = 0,5 + 0,3944 = 0,8944.$$

3) Події «першу страву замовили не менше 75 студентів» та «першу страву замовили не більше 74 студентів» протилежні, тому

$$P_{100}(0;74) = 1 - P_{100}(75;100) = 1 - 0,8944 = 0,1056.$$

1) Розрахунки проведено поетапно:

1.1)

B3 f_x =НОРМАЛИЗАЦИЯ(D1;B1*B2;КОРЕНЬ(B1*B2*D2))

	A	B	C	D	E	F	G
1	*n*	100	*k1*	75	*k2*	90	
2	*p*	0,8	*q*	0,2			
3	*x1*	-1,25					

1.2)

D3 f_x =НОРМАЛИЗАЦИЯ(F1;B1*B2;КОРЕНЬ(B1*B2*D2))

	A	B	C	D	E	F	G
1	*n*	100	*k1*	75	*k2*	90	
2	*p*	0,8	*q*	0,2			
3	*x1*	-1,25	*x2*	2,5			

1.3)

B4 f_x =НОРМСТРАСП(B3)-0,5

	A	B	C	D	E	F
1	*n*	100	*k1*	75	*k2*	90
2	*p*	0,8	*q*	0,2		
3	*x1*	-1,25	*x2*	2,5		
4	$\Phi(x_1)$	-0,3944				

1.4)

D4 | f_x =НОРМСТРАСП(D3)-0,5

	A	B	C	D	E	F
1	n	100	$k1$	75	$k2$	90
2	p	0,8	q	0,2		
3	$x1$	-1,25	$x2$	2,5		
4	$\Phi(x_1)$	-0,3944	$\Phi(x_2)$	0,4938		

1.5)

D5 | f_x =D4-B4

	A	B	C	D	E	F
1	n	100	$k1$	75	$k2$	90
2	p	0,8	q	0,2		
3	$x1$	-1,25	$x2$	2,5		
4	$\Phi(x_1)$	-0,3944	$\Phi(x_2)$	0,4938		
5		$\Phi(x_1)-$	$\Phi(x_2)$	0,8881		

2) Розрафунки проводяться аналогічно попередньому прикладу, нижче приведено остан ній крок розрахунків:

D5 | f_x =D4-B4

	A	B	C	D	E	F
1	n	100	$k1$	75	$k2$	100
2	p	0,8	q	0,2		
3	$x1$	-1,25	$x2$	5		
4	$\Phi(x_1)$	-0,3944	$\Phi(x_2)$	0,5		
5		$\Phi(x_2)-\Phi(x_1)$		0,8943		

3) Розраховується на основі 2):

D6 | f_x =1-D5

	A	B	C	D	E	F
1	n	100	$k1$	75	$k2$	100
2	p	0,8	q	0,2		
3	$x1$	-1,25	$x2$	5		
4	$\Phi(x_1)$	-0,3944	$\Phi(x_2)$	0,5		
5		$\Phi(x_2)-\Phi(x_1)$		0,8943		
6		$1-(\Phi(x_2)-\Phi(x_1))$		0,1057		

1)

$n := 100 \quad k1 := 75 \quad k2 := 90$

$p := 0.8 \qquad q := 1 - p = 0.2$

$x1 := \dfrac{k1 - n \cdot p}{\sqrt{n \cdot p \cdot q}} = -1.25 \qquad x2 := \dfrac{k2 - n \cdot p}{\sqrt{n \cdot p \cdot q}} = 2.5$

$\Phi1 := \mathrm{pnorm}(x1, 0, 1) - 0.5 = -0.394$

$\Phi2 := \mathrm{pnorm}(x2, 0, 1) - 0.5 = 0.494$

або

$\Phi1 := \mathrm{cnorm}(x1) - 0.5 = -0.394$

$\Phi2 := \mathrm{cnorm}(x2) - 0.5 = 0.494$

$P := \Phi2 - \Phi1 = 0.888$

2)

$n := 100 \quad k1 := 75 \quad k2 := 100$

$p := 0.8 \qquad q := 1 - p = 0.2$

$x1 := \dfrac{k1 - n \cdot p}{\sqrt{n \cdot p \cdot q}} = -1.25 \qquad x2 := \dfrac{k2 - n \cdot p}{\sqrt{n \cdot p \cdot q}} = 5$

$\Phi1 := \mathrm{pnorm}(x1, 0, 1) - 0.5 = -0.394$

$\Phi2 := \mathrm{pnorm}(x2, 0, 1) - 0.5 = 0.5$

$P := \Phi2 - \Phi1 = 0.894$

3) Розраховується на основі 2):

$1 - P = 0.106$

Відповідь: 1) ймовірність того, що першу страву замовлять не менше 75 і не більше 90 студентів дорівнює 0,8882; 2) ймовірність того, що першу страву замовлять не менше 75 студентів дорівнює 0,8944; 3) ймовірність того, що першу страву замовлять не більше 74 студентів дорівнює 0,1056.

Нехай проведено n незалежних випробувань за схемою Бернуллі, в яких подія A відбулась m разів. Якщо кількість випробувань n велика, то за допомогою інтегральної теореми Муавра-Лапласа можна оцінити ймовірність відхилення відносної частоти $W(A)=\dfrac{m}{n}$ появи події від ймовірності p її появи в одному випробуванні.

<u>Наслідок.</u> Ймовірність того, що при проведенні n незалежних випробувань відхилення відносної частоти події А від її ймовірності за модулем не перевищує ε, визначається за формулою

$$P\left(\left|\frac{m}{n}-p\right|\le\varepsilon\right)\approx 2\Phi\left(\varepsilon\cdot\sqrt{\frac{n}{pq}}\right). \tag{3.9}$$

Для обчислення аргументу функції Лапласа при відомому її значенні використовуються:

- статистична функція **НОРМОБР(*вероятность*;*среднее*;*стандартное_откл*)**, де *вероятность* – відповідне значення ймовірності, *среднее* – середнє арифметичне значення розподілу, *стандартное_откл* – стандартне відхилення розподілу, причому *среднее*=0, *стандартное_откл*=1; також можна використати статистичну функцію **НОРМСТОБР(*вероятность*)**, де *вероятность* – відповідне значення ймовірності, причому замість *вероятность* слід підставляти значення збільшене на 0,5;
- функція **qnorm(*p*,**μ **,**σ**)**, де p – рівень квантилі розподілу, μ – середнє арифметичне значення розподілу, σ – стандартне відхилення розподілу, причому $\mu=0, \sigma=1,$ замість p потрібно підставляти значення збільшене на 0,5.

<u>Приклад 3.6.</u> Відділ контролю якості продукції перевіряє партію кетчупу розфасованого у банки. Ймовірність того, що за якимось із критеріїв банка кетчупу буде забракована дорівнює 0,05. Скільки потрібно перевірити банок, щоб із ймовірністю 0,99 за абсолютною величиною відносна частота появи забракованих банок відхилялась від вказаного відсотку браку не більше, ніж на 0,01?

<u>Розв'язання.</u> У цій задачі потрібно скористатись інтегральною теоремою Муавра-Лапласа. За умовою $p=0,05$; $q=1-0,05=0,95$; $\varepsilon=0,01$; $P\left(\left|\frac{m}{n}-0,05\right|\le 0,01\right)=0,99$.

За формулою (3.9): $P\left(\left|\frac{m}{n}-0,05\right|\le 0,01\right)\approx 2\Phi\left(0,01\cdot\sqrt{\frac{n}{0,05\cdot 0,95}}\right)=0,99$

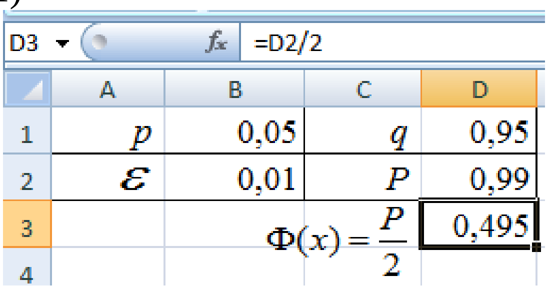
або $\Phi\left(0,01\cdot\sqrt{\frac{n}{0,05\cdot 0,95}}\right)=\frac{0,99}{2}$, або $\Phi\left(0,01\cdot\sqrt{\frac{n}{0,0475}}\right)=0,495$.

За таблицею (додаток 2): $0,495=\Phi(2,576)$, тому $0,01\sqrt{\frac{n}{0,0475}}=2,576$.

Звідси $n=\left(\frac{2,576}{0,01}\right)^2\cdot 0,0475\approx 3152$.

Розрахунки проведено поетапно:

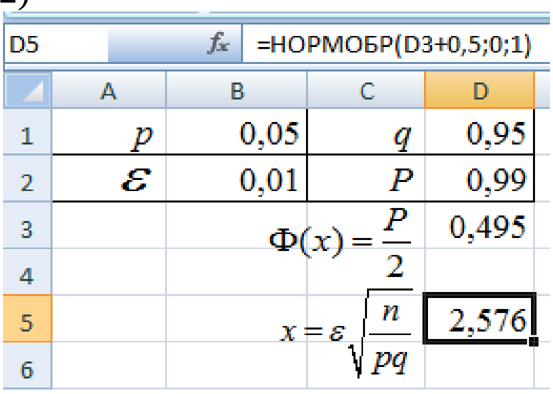
1)



D3 =D2/2

	A	B	C	D
1	p	0,05	q	0,95
2	ε	0,01	P	0,99
3			$\Phi(x)=\frac{P}{2}$	0,495
4				

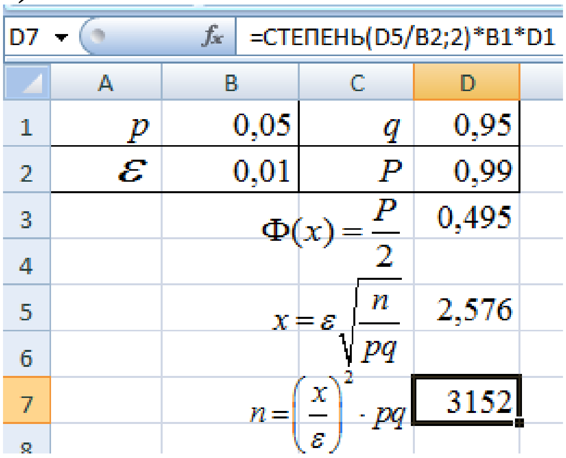
2)



D5 =НОРМОБР(D3+0,5;0;1)

	A	B	C	D
1	p	0,05	q	0,95
2	ε	0,01	P	0,99
3			$\Phi(x)=\frac{P}{2}$	0,495
4				
5			$x=\varepsilon\sqrt{\frac{n}{pq}}$	2,576
6				

3)



D7 f_x =СТЕПЕНЬ(D5/B2;2)*B1*D1

	A	B	C	D
1	p	0,05	q	0,95
2	ε	0,01	P	0,99
3			$\Phi(x)=\dfrac{P}{2}$	0,495
4				
5			$x=\varepsilon\sqrt{\dfrac{n}{pq}}$	2,576
6				
7			$n=\left(\dfrac{x}{\varepsilon}\right)^2\cdot pq$	3152

Примітка. У розрахунках була використана математична функція **СТЕПЕНЬ(*число*;*степень*)**, за допомогою якої можна піднести задане число у заданий степінь.

$$p := 0.05 \qquad q := 1 - p = 0.95 \qquad \varepsilon := 0.01$$

$$P := 0.99 \qquad P1 := \frac{P}{2} + 0.5$$

$$\mathrm{qnorm}(P1, 0, 1) = 2.576$$

$$n := 0$$

Given

$$\varepsilon\cdot\sqrt{\frac{n}{p\cdot q}} = 2.576$$

$$\mathrm{Find}(n) = 3.152\times 10^{3}$$

Примітка. У розрахунках було використано блок розв'язання рівняння **Given-Find**, який застосовується для розв'язання лінійних і нелінійних рівнянь та систем рівнянь. Функція **Find(**$x_1, x_2, \ldots$**)**, повертає точний розв'язок рівняння. Слід пам'ятати, що для невідомого (невідомих) потрібно задати початкове наближення (у приведеному прикладі $n := 0$); для задання рівнянь використовується « **=** », яке міститься на підпанелі **Boolean** (Булевы операторы) панелі **Math** (Математика).

Відповідь: потрібно перевірити 3152 банки.

Практичне заняття 3

3.1. Для вибору етикеток на новий вид продукції фірми «Колобок» було проведене опитування, яке показало, що 70 % опитаних вибрали один вид етикетки. Яка ймовірність того, що серед 7 навмання вибраних опитаних, рівно 4 виберуть той самий вид етикетки?

3.2. На екзамен виноситься 5 дисциплін. Студент оцінив свої шанси наступним чином: якщо готуватись до 5 дисциплін, то ймовірність скласти екзамен по кожній з них становитиме 60 %, до 4 – 70 %, до 3 – 85%. Потрібно здати екзамени щонайменше по 3 дисциплінам. Який із шляхів обрати студенту, щоб як найкраще досягти мети?

3.3. Ймовірність того, що бухгалтер не допустить помилку при заповненні одного звіта складає 0,9. Знайти ймовірність того, що із 9 звітів, зданих бухгалтером, не більше 2 – будуть містити помилки.

3.4. Туристична компанія «Сокіл» кожного літа відправляє туристів у країни Європи. Практика показує, що ймовірність замовлення туру в Іспанію становить 0,6. Знайти ймовірність того, що із 8 клієнтів компанії 4 замовлять тур в Іспанію.

3.5. Ймовірність продати кожний із 4 пакетів акцій у період їх обвалу за оптимальною ціною становить 0,35. Знайти ймовірність продажу: 1) усіх пакетів акцій за оптимальною ціною; 2) хоча б 2 пакетів акцій.

3.6. Для деякої місцевості ймовірність народження хлопчика дорівнює 0,51. У сім'ї, яка проживає у цій місцевості, 4 дітей. Яка ймовірність того, що у цій сім'ї: 1) усі хлопчики; 2) усі дівчатка; 3) хоча б 1 хлопчик; 4) хоча б 1 дівчинка; 5) більше дівчаток, ніж хлопчиків.

3.7. На підприємстві працює 5 технологічних ліній по розливу соків. Ймовірність порушення роботи кожної з них становить 0,15. Знайти 1) ймовірність того, що не менше 3 ліній працюють з порушеннями; 2) найвірогідніше число ліній, які працюють з порушеннями, та його ймовірність.

3.8. Іван-царевич випускає 7 стріл. Ймовірність того, що кожна із цих стріл потрапить до Царівни-жаби становить 0,35. Знайти найвірогідніше число стріл, які потраплять до Царівни-жаби.

3.9. У лотереї випущено 900 білетів. Ймовірність того, що кожен із них виграшний, становить 0,15. Знайти найвірогідніше число виграшних білетів та його ймовірність.

3.10. Потік економічного факультету складає 79 студентів. Ймовірність того, що студент не відвідає лекцію з теорії ймовірностей становить 0,05. Знайти найвірогідніше число студентів, які відвідають лекцію.

3.11. На олійно-жирове підприємство «Сонях» для проходження практики було відправлено 19 студентів. Ймовірність того, що за результатами проходження практики кожному з них запропонують роботу становить 0,8. Знайти найвірогідніше число студентів, яким запропонують роботу після успішного проходження практики.

3.12. Відомо, що 70 % випускників університету успішно знаходять роботу. Яким був обсяг випуску університету у поточному році, якщо успішно знайшли роботу 770 випускників?

3.13. Маркетологами підприємства було оцінено, що ймовірність реалізації кожної із пляшок води нового виду становитиме 0,75. Скільки пляшок води нового виду потрібно підготувати для продажу, щоб найвірогідніше число реалізованих пляшок становило 750?

3.14. Оцінити ймовірність реалізації 1500 нових товарів, випуск яких освоїло підприємство, якщо відомо, що із 2200 – реалізовано було 1980.

3.15. На потоці вчаться 178 студентів. Яка ймовірність того, що у трьох з них день народження випадає на 1 вересня?

3.16. З практики відомо, що 0,1 % пакетів соку мають порушення герметичності. Обчислити ймовірність того, що серед 500 пакетів з порушеннями герметичності виявляться: 1) 2 пакети; 2) хоча б 1 пакет.

3.17. Ймовірність того, що комплект документів на отримання пільг по оплаті житла заповнений невірно, становить 0,001. Знайти ймовірність того, що із 2000 комплектів документів невірно оформленими будуть 1) 5; 2) менше 3; 3) хоча б 1.

3.18. На одній із технологічних ліній підприємства масло розфасовується у пачки. Відомо, що на кожні 1000 пачок 3 не відповідають стандарту. Визначити ймовірність того, що із 500 відправлених на реалізацію пачок не відповідатимуть стандарту 1) 2; 2) хоча б 2.

3.19. Бібліотека університету налічує 2000 книжок нового надходження. Ймовірність того, що кожна із цих книжок має дефект 0,002. Знайти ймовірність того, що із книжок нового надходження мають дефект: 1) не менше 3; 2) 5.

3.20. Визначити ймовірність відмови 50 агрегатів із 130 працюючих, якщо ймовірність відмови агрегату дорівнює 0,12.

3.21. На підприємстві працює 135 співробітників. Ймовірність звільнення за власним бажанням кожного з них за певний період становить 0,25. Знайти ймовірність того, що за певний період звільняться за власним бажанням 28 співробітників.

3.22. Відділ контролю якості обслуговування клієнтів банку «Стогав» проводить опитування. Практика показує, що до 65 % клієнтів можна додзвонитись з першого разу. Яка ймовірність того, що із 500 клієнтів до 350 оператор додзвониться з першого разу?

3.23. Для вивчення попиту на питний йогурт маркетологи підприємства «Розливай» вирішили провести опитування 350 споживачів. З практики відомо, що відповідати на питання анкети погоджуються 75 % опитуваних. Знайти ймовірність того, що 310 споживачів погодяться відповісти на питання анкети.

3.24. Для підтвердження результату експерименту потрібно отримати 150 позитивних спроб. Випадки отримання позитивної і негативної спроб у результаті експерименту рівноймовірні. Яка ймовірність підтвердження експерименту, якщо було зроблено 700 спроб?

3.25. Книга, обсягом 500 сторінок, містить певну кількість помилок. Відомо, що ймовірність помилки на кожній із сторінок цієї книги дорівнює 0,2.

Знайти ймовірність того, що помилки містять: 1) 10 сторінок; 2) не більше 200 сторінок.

3.26. Ймовірність того, що відвідувач супермаркету придбає хліб становить 0,8. Знайти ймовірність того, що із 100 відвідувачів хліб придбають 1) не більше 49; 2) від 50 до 90; 3) не менше 50 відвідувачів.

3.27. Підприємство нараховує 60 співробітників. Ймовірність піти на лікарняний кожного з них протягом грудня дорівнює 0,45. Яка ймовірність того, що протягом грудня на лікарняний підуть не більше половини співробітників?

3.28. Ймовірність реалізувати певний виріб у період кризи становить 0,3. Знайти ймовірність того, що із 200 виробів реалізованими будуть: 1) від 120 до 150; 2) не менше 150; 3) не більше 100.

3.29. На обліку служби зайнятості району стоїть 580 безробітних. Відомо, що перша пропозиція по працевлаштуванню задовольняє тільки 40 %. Яка ймовірність того, що з першого разу будуть працевлаштованими: 1) не більше 30 осіб; 2) від 200 до 300 осіб.

3.30. Відомий письменник Незнайко вирішив написати роман у віршах. У його розпорядженні є 100 сторінок. Ймовірність того, що йому сподобається написане і він не розірве та не викине свій шедевр, для кожної із сторінок становить 0,15. Знайти ймовірність того, що роман матиме обсяг хоча б 30 сторінок.

3.31. Проводиться соціологічне опитування населення, для якого відібрано 750 чоловік. Відомо, що кожний з опитаних незалежно від інших може дати нещиру відповідь із ймовірністю 0,25. Знайти ймовірність того, що відносна частота появи нещирих відповідей опитуваних відхилиться від її ймовірності за абсолютною величиною не більше, ніж на 0,01.

3.32. При надходженні грошей у банк потрапляє в середньому 0,1 % фальшивих купюр різним номіналом. Визначити ймовірність того, що із 1000 купюр, які надійшли у банк, відхилення від вказаного відсотку фальшивих за абсолютною величиною не перевищує 0,02.

3.33. Ймовірність того, що кожному із 200 осіб, які звернулись у туристичну фірму «Земля», підберуть путівку дорівнює 0,75. Знайти ймовірність того, що відносна частота задоволених осіб відхилиться від її ймовірності за абсолютною величиною не більше, ніж на 0,015.

3.34. Скільки виробів потрібно перевірити для визначення відсотку браку, щоб із ймовірністю 0,99 можна було стверджувати, що відхилення частоти появи бракованих виробів від їх ймовірності, яка дорівнює 0,85, за абсолютною величиною не перевищувало 0,01?

3.35. Скільки разів потрібно кидати монету, щоб із ймовірністю 0,9544 стверджувати, що відносна частота випадання герба відхиляється від 0,5 не більше ніж на 0,05?

Лекція 4. Дискретні випадкові величини

4.1. Поняття випадкової величини

4.2. Закони розподілу дискретних випадкових величин

4.3. Дії над випадковими величинами

4.1. Поняття випадкової величини

Одним з основних понять теорії ймовірностей є поняття випадкової величини.

Раніше вже розглядалось, що поняття події було введене як абстрактна модель якісної ознаки, в якої є сенс лише альтернативного судження «є» ознака або її «немає». Для аналізу випадкових явищ важливим є введення нового поняття – випадкової величини, яка являє собою абстрактну модель кількісної ознаки.

Випадкова величина характеризує кількісний результат стохастичного експерименту. Прикладами величин, які приймають різні числові значення під впливом різноманітних випадкових обставин, можуть служити:

1) число народжених дітей протягом доби в місті Києві;

2) число викликів, що надійшли на телефонну станцію протягом певного проміжку часу;

3) число влучень при п'яти пострілах;

4) число пострілів до першого влучення;

5) витрати електроенергії на підприємстві протягом місяця;

6) дальність польоту снаряду.

У наведених прикладах зустрічалися два типи випадкових величин: дискретні, множина значень яких скінченна (приклади 1, 2, 3) або нескінчена злічена (приклад 4); неперервні, множина значень яких належить певному числовому проміжку (приклади 5, 6).

Ці величини так чи інакше характеризували проведений стохастичний експеримент. Кожна з цих величин під впливом випадкових обставин може приймати різні значення. Заздалегідь вказати, яке значення набуде ця величина, неможливо, так як воно змінюється випадковим чином.

Теоретико-множинна трактовка основних понять теорії ймовірностей дозволяє дати наступне означення випадкової величини.

Означення. *Випадковою величиною* X називається функція, задана на множині елементарних подій (або у просторі елементарних подій), тобто $X = f(\omega)$.

Для *дискретної випадкової величини* множина можливих значень випадкової величини, тобто множина значень функції, скінчена або злічена, для неперервної – нескінчена і незлічена.

Приклад 4.1. Монету підкидають двічі. Описати значення випадкової величини X – числа появи герба на множині елементарних подій.

Розв'язання. Вкажемо множину елементарних подій. Позначимо через Г – появу «герба» в одному підкиданні, Р – появу «решки». Маємо множину елементарних подій: $\omega_1 = ГГ$; $\omega_2 = ГР$; $\omega_3 = РГ$; $\omega_4 = РР$. Тоді випадкова величина набуде значення $X(\omega_1) = 2$; $X(\omega_2) = 1$; $X(\omega_3) = 1$; $X(\omega_4) = 0$. Отже, $X = \{0; 1; 2\}$.

Відповідь: $X = \{0; 1; 2\}$.

Випадкові величини будемо позначати великими літерами латинського алфавіту $X, Y, Z, \ldots$, а їх значення відповідно маленькими літерами $x, y, z, \ldots$

Зауваження. Випадкову подію можна розглядати як частинний випадок випадкової величини, яка приймає два значення: 1, якщо подія відбулася, і 0, якщо ні.

4.2. Закони розподілу дискретних випадкових величин

Для завдання випадкової величини необхідно вказати не тільки множину її можливих значень (що було б достатньо при описі звичайної змінної величини), але і вказати ймовірності, з якими набуваються ті або інші значення. Таке повне завдання дає закон розподілу.

<u>*Означення*</u>. *Законом розподілу випадкової величини* називається будь-яке співвідношення, яке встановлює зв'язок між можливими значеннями випадкової величини і відповідними їм ймовірностями.

Про випадкову величину кажуть, що вона «розподілена» за даним законом розподілу, або «підпорядкована» цьому закону розподілу.

Для дискретної випадкової величини закон розподілу може бути задано у вигляді таблиці, графічно і аналітично (за допомогою формули).

Нехай задано дискретну випадкову величину X, яка може набувати значення $x_1, x_2, ..., x_i, ..., x_n$ з ймовірностями $P(X = x_1) = p_1$, $P(X = x_2) = p_2$, … $P(X = x_i) = p_i$, … $P(X = x_n) = p_n$. Відповідність між двома послідовностями $\{x_i\}$ і $\{p_i\}$ може бути у цьому випадку задана у вигляді таблиці, в якій прийнято x_i розташовувати в порядку зростання:

X	x_1	x_2	$\cdots$	x_i	$\cdots$	x_n
P	p_1	p_2	$\cdots$	p_i	$\cdots$	p_n

Така таблиця називається ще *рядом розподілу дискретної випадкової величини*.

Події $X = x_1$, $X = x_2$, …, $X = x_n$ в сукупності утворюють повну групу несумісних подій, тому сума їх ймовірностей дорівнює одиниці. Отже, для розподілу будь-якої дискретної випадкової величини

$$\sum_{i=1}^{n} P(X = x_i) = \sum_{i=1}^{n} p_i = 1. \qquad (4.1)$$

Таблицю розподілу можна зобразити графічно, якщо по осі абсцис відкладати значення випадкової величини, а по осі ординат – відповідні їм ймовірності. З'єднання отриманих точок з координатами $(x_k; p_k)$ утворює ламану, яка називається многокутником розподілу або полігоном розподілу ймовірностей (рис. 4.1).

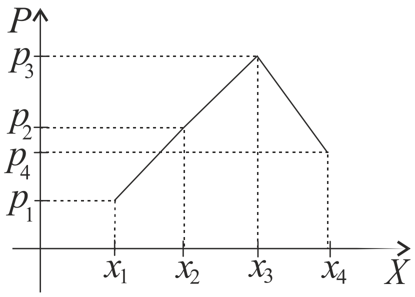


Рис. 4.1

Означення. Значення випадкової величини, ймовірність якої найбільша, називається *модою*.

На малюнку модою є x_3.

При аналітичному способі задання закону розподілу дискретної випадкової величини ймовірність відповідного значення шукається за формулою

$$p_k = P(X = x_k) = f(x_k),\ k = 1, 2 \dots$$

Наприклад: $P(X = k) = \dfrac{\lambda^k}{k!} e^{-\lambda},\ k = 0, 1, \dots$

Приклад 4.2. Ймовірність того, що студент складе семестровий екзамен у сесію з дисциплін A і B дорівнює відповідно 0,8 і 0,9. Записати закон розподілу числа семестрових екзаменів, які може скласти студент в сесію.

Розв'язання. Випадкова величина X – число семестрових екзаменів, які може скласти студент в сесію, набуває значень $X = \{0; 1; 2\}$.

Позначимо через A_1 – подію, що студент склав перший екзамен, A_2 – склав другий екзамен. Через $\overline{A}_1$, $\overline{A}_2$ – відповідно не склав екзамен. Тоді

$$P(X = 0) = P(\overline{A}_1 \overline{A}_2) = P(\overline{A}_1) P(\overline{A}_2) = (1 - 0,8)(1 - 0,9) = 0,02;$$

$$P(X = 1) = P(A_1 \overline{A}_2) + P(\overline{A}_1 A_2) = P(A_1) P(\overline{A}_2) + P(\overline{A}_1) P(A_2) =$$
$$= 0,8 \cdot 0,1 + 0,9 \cdot 0,2 = 0,26;$$

$$P(X=2)=P(A_1A_2)=P(A_1)P(A_2)=0,72.$$

Отже, закон розподілу має вигляд:

X	0	1	2
P	0,02	0,26	0,72

$$\sum_{i=1}^{3} p_i = 0,02+0,26+0,72=1.$$

Відповідь:

X	0	1	2
P	0,02	0,26	0,72

Приклад 4.3. Скласти закон розподілу величини виграшу на один лотерейний білет, якщо на кожні 100 білетів розігрується 1 виграш в 1000 грн., 10 – по 100 грн., 25 – по 10 грн.

Розв'язання. Випадкова величина X – величина виграшу на один лотерейний білет, може набувати значення $X=\{0;10;100;1000\}$.

$$p_1=P(X=0)=\frac{64}{100};\ \ p_2=P(X=10)=\frac{25}{100};$$

$$p_3=P(X=100)=\frac{10}{100};\ \ p_4=P(X=1000)=\frac{1}{100}.$$

Отже, закон розподілу має вигляд:

X	0	10	100	1000
P	0,64	0,25	0,1	0,01

$$\sum_{i=1}^{4} p_i = 0,64+0,25+0,1+0,01=1.$$

Відповідь:

X	0	10	100	1000
P	$0,64$	$0,25$	$0,1$	$0,01$

Розглянемо закони розподілу дискретних випадкових величин, які часто зустрічаються в різноманітних моделях.

4.2.1. Рівномірний закон розподілу на множині $\{1, 2, ..., n\}$

Випадкова величина X, що приймає цілочисельні значення від 1 до n має рівномірний розподіл, якщо

$$P\left(X = m\right) = \frac{1}{n}, \ m = 1, 2 \ldots n.$$

Многокутником розподілу випадкової величини X є відрізок прямої. Кінці цього відрізку мають координати $\left(1; \frac{1}{n}\right)$ і $\left(n; \frac{1}{n}\right)$ (рис. 4.2).

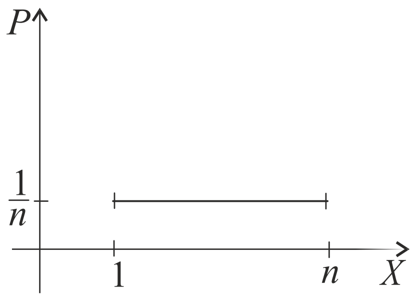


Рис. 4.2

4.2.2. Гіпергеометричний закон розподілу

Випадкова величина X має гіпергеометричний розподіл, якщо

$$P(X = m) = \frac{C_M^m C_{N-M}^{n-m}}{C_N^n}, \ m = 0, 1 \ldots \min\left(n, M\right).$$

<u>Приклад 4.4.</u> В партії з 10 виробів 6 виробів першого сорту, а інші – другого сорту. З партії беруть 3 вироби. Скласти закон розподілу числа виробів першого сорту серед відібраних виробів.

Розв’язання. Випадкова величина X – число виробів першого сорту серед відібраних, може приймати значення $x_1 = 0;\ x_2 = 1;\ x_3 = 2;\ x_4 = 3.$

$N = 10;\ n = 3;\ M = 6.$

$$p_1 = P(X = 0) = \frac{C_6^0 C_4^3}{C_{10}^3} = \frac{4}{120};\ p_2 = P(X = 1) = \frac{C_6^1 C_4^2}{C_{10}^3} = \frac{36}{120};$$

$$p_3 = P(X = 2) = \frac{C_6^2 C_4^1}{C_{10}^3} = \frac{60}{120};\ p_4 = P(X = 3) = \frac{C_6^3 C_4^0}{C_{10}^3} = \frac{20}{120}.$$

Заповнимо таблицю розподілу:

X	0	1	2	3
P	$\frac{4}{120}$	$\frac{36}{120}$	$\frac{60}{120}$	$\frac{20}{120}$

$$\sum_{i=1}^{4} p_i = \frac{4}{120} + \frac{36}{120} + \frac{60}{120} + \frac{20}{120} = 1.$$

Відповідь:

X	0	1	2	3
P	$\frac{4}{120}$	$\frac{36}{120}$	$\frac{60}{120}$	$\frac{20}{120}$

4.2.3. Геометричний закон розподілу

Випадкова величина X має геометричний розподіл, якщо

$$p_m = P(X = m) = q^{m-1} p, \text{ де } m = 1, 2\ldots,\ 0 < p < 1,\ q = 1 - p.$$

Просумувавши нескінченно спадну геометричну послідовність, маємо:

$$\sum_{m=1}^{\infty} p_m = \sum_{m=1}^{\infty} pq^{m-1} = p\frac{1}{1-q} = 1.$$

Геометричний розподіл має випадкова величина $X,$ яка дорівнює числу випробувань у схемі Бернуллі до першої появи події, якщо ймовірність появи події в одному випробуванні $p.$

4.2.4. Закон розподілу Пуассона

Випадкова величина X має розподіл Пуассона з параметром λ, якщо

$$p_m = P(X = m) = \frac{\lambda^m}{m!} e^{-\lambda}, \; m = 0, 1, 2 \ldots, \; \lambda > 0.$$

Покажемо, що $\sum_{m=0}^{\infty} p_m = 1.$

Дійсно, $\sum_{m=0}^{\infty} p_m = \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} e^{-\lambda} = e^{-\lambda} \sum_{m=0}^{\infty} \frac{\lambda^m}{m!} = e^{-\lambda} \cdot e^{\lambda} = 1.$

4.2.5. Біноміальний закон розподілу

Дискретна випадкова величина X, яка приймає значення від 0 до n має біноміальний розподіл, якщо $P\left(X = m\right) = p_n\left(m\right) = C_n^m p^m q^{n-m}$.

Біноміальний розподіл має випадкова величина X - число появ події A у схемі Бернуллі, якщо ймовірність появи події A в одному випробуванні p, а непояви $q = 1 - p$.

4.3. Дії над випадковими величинами

Розглянемо поняття незалежних випадкових величин.

Означення. Дві випадкові величини називаються *незалежними*, якщо закон розподілу однієї з них не змінюється від того, яких можливих значень набуває інша.

Отже, якщо дискретна випадкова величина X може набути значення x_i, $i = \overline{1, n}$, а випадкова величина Y – значення y_j, $j = \overline{1, m}$, то незалежність дискретних випадкових величин X і Y означає незалежність подій $X = x_i$, $Y = y_j$, $i = \overline{1, n}$; $j = \overline{1, m}$. В іншому випадку величини називаються залежними.

Наприклад, якщо є білети двох різних грошових лотерей, то випадкові величини X і Y, що є відповідно виграшем по кожному білету (в грошових одиницях) незалежні.

Визначимо математичні операції з дискретними випадковими величинами.

Нехай задані дві випадкові величини X і Y з законами розподілу відповідно

X	x_1	x_2	$\cdots$	x_n
p_i	p_1	p_2	$\cdots$	p_n

Y	y_1	y_2	$\cdots$	y_m
p_j	p_1	p_2	$\cdots$	p_m

1) Добутком CX випадкової величини X на сталу величину C називається випадкова величина Cx_i, $i=\overline{1,n}$ з тими ж ймовірностями p_i, $i=\overline{1,n}$.

2) m-м степенем випадкової величини X, тобто X^m називається випадкова величина, яка набуде значення x_i^m $i=\overline{1,n}$, з тими ж ймовірностями p_i, $i=\overline{1,n}$.

<u>**Приклад 4.5.**</u> Задана дискретна випадкова величина X:

X	-3	-2	1	4
p	$0,2$	$0,3$	$0,4$	$0,1$

Знайти закон розподілу дискретної випадкової величини: а) $Y=2X$; б) $Z=X^2$

<u>Розв’язання.</u>

а) Враховуючи 1), $y_1=2\cdot(-3)=-6;\ P(Y=-6)=0,2$ і т.д.:

Y	-6	-4	2	8
p	$0,2$	$0,3$	$0,4$	$0,1$

б) Враховуючи 2), при $m=2$, $z_1=(-3)^2=9$; $P(Z=9)=0,2$ і т.д.:

Z	9	4	1	16
p	$0,2$	$0,3$	$0,4$	$0,1$

Відповідь: а)

Y	–6	–4	2	8
p	$0,2$	$0,3$	$0,4$	$0,1$

б)

Z	9	4	1	16
p	$0,2$	$0,3$	$0,4$	$0,1$

3) Сумою (різницею або добутком) випадкових величин X і Y називається випадкова величина, яка набуває всі можливі значення виду x_i+y_j (x_i-y_j або $x_i y_j$), де $i=\overline{1,n}$; $j=\overline{1,m}$ з ймовірностями p_{ij} того, що випадкова величина X набуває значення x_i, а Y – значення y_j, $p_{ij}=P\left[(X=x_i)(Y=y_j)\right]$.

Якщо випадкові величини X та Y незалежні, тобто незалежні будь-які події $X=x_i$; $Y=y_j$, то за теоремою множення ймовірностей незалежних випадкових подій $p_{ij}=P(X=x_i)P(Y=y_j)=p_i p_j$.

Приклад 4.6. Задано закони розподілу двох незалежних дискретних випадкових величин:

X	0	3	4
p_i	$0,5$	$0,2$	$0,3$

Y	–2	1	3
p_j	$0,2$	$0,5$	$0,3$

Знайти закон розподілу дискретної випадкової величини: а) $Z = X - Y$; б) $U = XY$.

<u>Розв’язання.</u> а) Для зручності знаходження всіх значень та їх ймовірностей потрібно скласти допоміжну таблицю:

	y_j	-2	1	3
x_i	p_j \ p_i	0,2	0,5	0,3
0	0,5	2 / 0,1	-1 / 0,25	-3 / 0,15
2	0,2	4 / 0,04	1 / 0,1	-1 / 0,06
4	0,3	6 / 0,06	3 / 0,15	1 / 0,09

Так серед дев’яти значень Z є однакові, то відповідні ймовірності додаються за теоремою додавання ймовірностей несумісних подій.

Наприклад, значення $Z = X - Y = 1$ можна одержати, коли $X = 2$; $Y = 1$ (з ймовірністю $0,01$) або $X = 4$; $Y = 3$ (з ймовірністю $0,09$). Тоді $Z = 1$ набуває з ймовірністю 0,19. Тобто $P(Z = 1) = 0,1 + 0,09 = 0,19$.

Отже, закон розподілу Z має вигляд:

Z	-3	-1	1	2	3	4	6
p_k	0,15	0,31	0,19	0,1	0,15	0,04	0,06

$$\sum_{k=1}^{7} p_k = 0,15 + 0,31 + 0,19 + 0,1 + 0,15 + 0,04 + 0,06 = 1.$$

б) Розподіл $U = XY$ знаходимо аналогічно:

U	-8	-4	0	2	4	6	12
p_k	0,06	0,04	0,5	0,1	0,09	0,06	0,15

Відповідь:

а)

Z	–3	–1	1	2	3	4	6
p_k	0,15	0,31	0,19	0,1	0,15	0,04	0,06

б)

U	–8	–4	0	2	4	6	12
p_k	0,15	0,04	0,5	0,1	0,09	0,06	0,15

Практичне заняття 4

4.1. У партії із 15 виробів 13 відповідають стандарту. Навмання відібрано 2 вироби. Скласти закон розподілу дискретної випадкової величини X – кількості стандартних виробів серед відібраних.

4.2. Бухгалтер здає у податкову 15 звітів, причому 5 з них мають недоліки. навмання 4 звіти. Податковий інспектор навмання вибирає 4 звіти. Скласти закон розподілу дискретної випадкової величини X – кількості звітів із відібраних, які мають недоліки.

4.3. Для проходження переатестації працівнику було запропоновано вивчити 30 питань. Не маючи багато вільного часу, він вивчив тільки 20. У білет ввійшло 3 питання. Скласти закон розподілу дискретної випадкової величини X – кількості питань, на які працівник знав відповідь.

4.4. У ящику лежать 7 яблук і 5 груш. Навмання взято 4 фрукта. Скласти закон розподілу дискретної випадкової величини X – кількості яблук серед узятих фруктів.

4.5. Із коробки, у якій лежать 3 білих і 5 червоних куль, витягують без повернення одна за одною кулі до появи білої кулі. Скласти закон розподілу дискретної випадкової величини X – кількості витягнутих із урни куль.

4.6. Покупець шукає потрібний йому товар у магазинах до тих пір, поки не знайде. Ймовірність того, що він знайде товар у магазині дорівнює 0,8. Скласти закон розподілу дискретної випадкової величини X – кількості магазинів, які відвідає покупець.

4.7. Артем телефонує своїм однокурсникам до тих пір, поки не дізнається домашнє завдання з теорії ймовірностей. Ймовірність того, що він дізнається домашнє завдання становить 0,6. Скласти закон розподілу дискретної випадкової величини X – кількості однокурсників, яким зателефонує Артем.

4.8. У кожному із двох таймів футбольного матчу обидві команди разом забивають 3 м’ячі з ймовірністю 0,15; 2 м’ячі – з ймовірністю 0,2; 1 м’яч – з ймовірністю 0,35; не забивають жодного м’яча – з ймовірністю 0,3. Скласти закон розподілу дискретної випадкової величини X – кількості забитих у матчі м’ячів.

4.9. Для прийняття участі у конференції відібрано 3 студентів. Ймовірність того, що перший студент прийме участь у конференції дорівнює 0,6; другий – 0,7; третій – 0,8. Скласти закон розподілу дискретної випадкової величини X – кількості студентів, які прийматимуть учать у конференції.

4.10. Для проходження практики на підприємстві відібрано 4 студентів. Ймовірність того, що перший студент пройде практику успішно становить 0,8; другий – 0,9; третій – 0,7; четвертий – 0,6. Скласти закон розподілу дискретної випадкової величини X – кількості студентів, які пройдуть успішно практику на підприємстві.

4.11. На підвищення кваліфікації підприємство відправляє за кордон 3 працівників. Ймовірність того, що кожен з них володіє іноземною мовою становить 0,75. Скласти закон розподілу дискретної випадкової величини X – кількості працівників, які володіють іноземною мовою.

4.12. Партія виробів містить 8 % браку. Скласти закон розподілу дискретної випадкової величини X – кількості виробів із 3 відібраних, які містять брак.

4.13. Комплекс по виробництву та фасуванню морозива складається із 4 незалежно працюючих ліній. Ймовірність виходу з ладу кожної лінії протягом місяця дорівнює 0,1. Скласти закон розподілу дискретної випадкової величини X – кількості ліній, що вийшли з ладу протягом місяця.

4.14. На шляху автомобіля 5 світлофорів. Ймовірність того, що автомобіль не зупиниться на світлофорі дорівнює 0,5. Скласти закон розподілу дискретної випадкової величини X – кількості, які проїхав автомобіль без зупинки.

4.15. Ймовірність того, що банкомат при введенні пінкоду спрацює правильно складає 0,99. Клієнт має 3 спроби. Скласти закон розподілу дискретної випадкової величини X – кількості спроб введення клієнтом пінкоду у банкомат.

4.16. Студенту потрібно відшукати книгу в одній із 3 бібліотек. Ймовірність наявності книги у кожній із бібліотек дорівнює 0,7. Скласти закон розподілу дискретної випадкової величини X – кількості бібліотек, які відвідає студент для пошуку потрібної книги.

4.17. Дискретна випадкова величина X задана законом розподілу:

X	1	2	4
p	0,1	0,5	0,4

Знайти закон розподілу дискретної випадкової величини: а) $Y = 2X$; б) $Z = X^2$.

4.18. Дискретна випадкова величина X задана законом розподілу:

X	-1	0	3
p	0,2	0,3	0,5

Знайти закон розподілу дискретної випадкової величини: а) $Y = 5X$ б) $Z = X^2$.

4.19. Дискретна випадкова величина X задана законом розподілу:

X	-2	-1	2	5
p	0,2	0,4	0,3	0,1

Знайти закон розподілу дискретної випадкової величини: а) $Y = 4X$ б) $Z = X^2$.

4.20. Дискретна випадкова величина X задана законом розподілу:

X	-3	0	1	2
p	0,1	0,5	0,2	0,2

Знайти закон розподілу дискретної випадкової величини: а) $Y = 3X$ б) $Z = X^2$.

4.21. Задано закони розподілу двох незалежних дискретних випадкових величин:

X	-1	1
p_i	0,3	0,7

Y	0	3
p_j	0,4	0,6

Знайти закон розподілу дискретної випадкової величини: а) $X + Y$; б) $X - 2Y$; в) XY.

4.22. Задано закони розподілу двох незалежних дискретних випадкових величин:

X	1	3
p_i	0,6	0,4

Y	-1	5
p_j	0,2	0,8

Знайти закон розподілу дискретної випадкової величини: а) $Z = X + Y$; б) $V = X - 3Y$; в) $U = XY$.

4.23. Задано закони розподілу двох незалежних дискретних випадкових величин:

X	-1	1	2
p_i	0,3	0,5	0,2

Y	0	1	3
p_j	0,4	0,3	0,3

Знайти закон розподілу дискретної випадкової величини: а) $Z = X + Y$; б) $V = 4X - Y$; в) $U = XY$.

4.24. Задано закони розподілу двох незалежних дискретних випадкових величин:

X	0	2	4
p_i	0,4	0,2	0,4

Y	-1	0	2
p_j	0,1	0,3	0,6

Знайти закон розподілу дискретної випадкової величини: а) $Z = X + Y$; б) $V = 2X - 3Y$; в) $U = XY$.

Лекція 5. Числові характеристики дискретної випадкової величини

Повну інформацію про випадкову величину дає закон розподілу випадкової величини. Проте при розв’язуванні багатьох задач теорії ймовірності достатньо знати лише кілька чисел, які певною мірою характеризують випадкову величину. Такі числа називають *числовими характеристиками випадкової величини*.

Найважливішими з числових характеристик є математичне сподівання, що характеризує середнє значення випадкової величини, і дисперсія, яка характеризує відхилення значення випадкової величини від математичного сподівання.

5.1. Математичне сподівання дискретної випадкової величини, його властивості

Означення. Математичним сподівання дискретної випадкової величини X, яка набуває значення $x_1, x_2, ..., x_n$ з ймовірностями відповідно рівними $p_1, p_2, ..., p_n$ називається число, яке позначається символом $M(X)$, і дорівнює сумі добутків усіх її значень на відповідні їм ймовірності:

$$M(X) = \sum_{k=1}^{n} x_k p_k. \qquad (5.1)$$

Зауваження. Якщо число можливих значень дискретної випадкової величини нескінченне $x_1, x_2, ..., x_n, ...,$ тоді математичним сподіванням такої випадкової величини називається сума ряда (якщо він абсолютно збіжний):

$$M(X) = \sum_{k=1}^{\infty} x_k p_k. \tag{5.2}$$

Властивості математичного сподівання.

1^0. Математичне сподівання сталої величини C дорівнює самій сталій:

$$M(C) = C..$$

Дійсно, сталу C можна розглянути як випадкову величину, яка може приймати тільки одне значення C з ймовірністю, рівною 1.

2^0. Математичне сподівання двох випадкових величин дорівнює сумі їх математичних сподівань.

Розглянемо дві дискретні випадкові величини X і Y, які приймають значення $x_1, x_2, ..., x_n$ і $y_1, y_2, ..., y_m$ відповідно. Позначимо через p_k ймовірність прийняття випадковою величиною X значення x_k, через q_l ймовірність прийняття випадковою величиною Y значення y_l, а через r_{kl} – ймовірність того, що випадкова величина X прийме значення x_k, а Y – значення y_l. Тоді за означенням математичного сподівання

$$M(X+Y) = \sum_{k=1}^{n}\sum_{l=1}^{m}(x_k + y_l)r_{kl} = \sum_{k=1}^{n} x_k \sum_{l=1}^{m} r_{kl} + \sum_{l=1}^{m} y_l \sum_{k=1}^{n} r_{kl}. \tag{5.3}$$

Так як за теоремою про повну ймовірність $\sum_{l=1}^{m} r_{kl}$ – це ймовірність події, яка полягає у тому, що випадкова величина X приймає значення x_k при умові, що випадкова величина Y приймає одне із своїх можливих значень (що вірогідно), то $\sum_{l=1}^{m} r_{kl} = p_k$. Аналогічно $\sum_{k=1}^{n} r_{kl} = q_l$. Отже, підставимо отримане в формулу (5.3) і остаточно матимемо

$$M(X+Y)=\sum_{k=1}^{n}x_k p_k+\sum_{l=1}^{m}y_l q_l=M(X)+M(Y).$$

<u>Наслідок.</u> Математичне сподівання суми скінченого числа випадкових величин дорівнює сумі їх математичних сподівань.

3^0. Математичне сподівання добутку незалежних випадкових величин дорівнює добутку їх математичних сподівань.

Використовуючи позначення пункту 2^0, так як випадкові величини X і Y незалежні, маємо, що сумісна поява події $x_k y_l$ має ймовірність $r_{kl}=p_k q_l$. Тому

$$M(XY)=\sum_{k=1}^{n}\sum_{l=1}^{m}x_k y_l p_k q_l=\sum_{k=1}^{n}x_k p_k+\sum_{l=1}^{m}y_l q_l=M(X)M(Y).$$

<u>Наслідок.</u> Постійний множник можна виносити за знак математичного сподівання.

Дійсно, $M(CX)=M(C)M(X)=CM(X)$.

4^0. Якщо всі значення випадкової величини збільшити (зменшити) на сталу C, то на цю сталу C збільшиться (зменшиться) математичне сподівання цієї випадкової величини:

$$M(X\pm C)=M(X)\pm M(C).$$

Дійсно, з властивостей 1^0 і 3^0 матимемо

$$M(X\pm C)=M(X)\pm M(C)=M(X)\pm C.$$

<u>Означення.</u> Відхиленням випадкової величини X від свого математичного сподівання $M(X)$ називається випадкова величина $X-M(X)$.

5^0. Математичне сподівання відхилення випадкової величини від її математичного сподівання дорівнює нулю:

$$M(X-M(X))=0.$$

З того що $M(M(X))=M(X)$, так як $M(X)$ – стала величина, з властивостей 2^0 і 3^0 одержимо:

$$M(X-M(X))=M(X)-M(M(X))=M(X)-M(X)=0.$$

<u>Приклад 5.1.</u> Знайти математичне сподівання випадкової величини $Z = 6X - 3Y$, якщо $M(X) = 2$ і $M(Y) = 4$.

<u>Розв’язання.</u> Використовуючи властивості математичного сподівання 2^0 і 3^0, матимемо $M(Z) = M(6X - 3Y) = 6M(X) - 3M(Y) = 12 - 12 = 0$.

<u>Відповідь:</u> $M(Z) = 0$.

Для обчислення математичного сподівання випадкової величини:

- використовується математична функція **СУММПРОИЗВ(*массив1*; *массив2*;…)**, яка повертає суму добутків діапазонів або масивів, де ***массив1*,*массив2*;…** – від 2 до 255 масивів, компоненти яких потрібно перемножити, а потім додати результати;

- дискретні випадкові величини можна задавати, як масиви (матриці, вектори), використовуючи підпанель **Matrix** (Матрица) панелі **Math** (Математика). Після натискання на піктограму з зображенням матриці (або натискання комбінації клавіш <ctrl>+<M>), з’являється запит, щодо розміру матриці. Слід пам’ятати, що за замовчуванням елементи матриці, задані з використанням нижнього індексу, у *MathCad* нумеруються з нульового елемента. Щоб змінити цей порядок потрібно змінити значення вбудованої змінної ORIGIN. Функція **rows(*A*)** повертає кількість рядків у матриці ***A***. Функція **cols(*A*)** повертає кількість стовпців у матриці ***A***. Математичне сподівання обчислюється безпосередньо за формулою.

<u>Приклад 5.2.</u> Задано закони розподілу двох випадкових величин X і Y:

X	-0,1	0,1
P	1/2	1/2

і

X	-100	100
P	1/2	1/2

Знайти їх математичні сподівання.

<u>Розв'язання.</u> $M(X)=\frac{1}{2}\cdot(-0{,}1)+\frac{1}{2}\cdot 0{,}1=0$; $M(Y)=\frac{1}{2}\cdot(-100)+\frac{1}{2}\cdot 100=0$.

B4 fx =СУММПРОИЗВ(A2:A3;B2:B3)

	A	B	C	D	E
1	*X*	*P*			
2	-0,1	0,5			
3	0,1	0,5			
4	*M(X)*	0			

B4 fx =СУММПРОИЗВ(A2:A3;B2:B3)

	A	B	C	D	E
1	*Y*	*P*			
2	-100	0,5			
3	100	0,5			
4	*M(Y)*	0			

$\mathrm{ORIGIN} \equiv 1$

$X := (-0.1 \quad 0.1)$

$P := (0.5 \quad 0.5)$

$$M := \sum_{i=1}^{\mathrm{cols}(X)} \left(X_{1,i}\cdot P_{1,i}\right) = 0$$

$\mathrm{ORIGIN} \equiv 1$

$Y := (-100 \quad 100)$

$P := (0.5 \quad 0.5)$

$$M := \sum_{i=1}^{\mathrm{cols}(Y)} \left(Y_{1,i}\cdot P_{1,i}\right) = 0$$

<u>Відповідь:</u> $M(X)=0,\ M(Y)=0.$

У цьому прикладі випадкові величини X і Y мають однакові математичні сподівання, але їх відхилення від математичного сподівання вже суттєво відрізняються.

Отже, тільки математичне сподівання не дає достатньої характеристики випадкової величини. Для більш повної інформації про випадкову величину вводять числову характеристику, яка дозволяє судити, як розсіюються значення випадкової величини навколо її математичного сподівання. До такої характеристики відноситься дисперсія – міра розсіювання.

5.2. Дисперсія дискретної випадкової величини, її властивості. Середнє квадратичне відхилення

<u>*Означення.*</u> *Дисперсією* дискретної випадкової величини X називається число, яке дорівнює математичному сподіванню квадрата відхилення цієї випадкової величини X від її математичного сподівання і позначається символом $D(X)$:

$$D(X) = M\left[(X - M(X))^2\right] \tag{5.4}$$

Якщо дискретна випадкова величина X набуває значення $x_1, x_2, ..., x_n$ з ймовірностями, відповідно рівними $p_1, p_2, ..., p_n$, то формулу (5.4) можна переписати у вигляді:

$$D(X) = \sum_{i=1}^{n} (x_i - M(X))^2 p_i. \tag{5.5}$$

Якщо дискретна випадкова величина набуває нескінчене число значень, то дисперсія дорівнює сумі ряду: $D(X) = \sum_{i=1}^{\infty} (x_i - M(X))^2 p_i$, при умові, що цей ряд збіжний.

Дисперсія випадкової величини як у *Excel*, так і у *MathCad* обчислюється безпосередньо за формулою, причому у *Excel* використовуються:

- крім математичної функції **СУММПРОИЗВ(*массив1*;*массив2*;…)**, описаної раніше, ще математична функція **СТЕПЕНЬ(*число*; *степень*)**, яка повертає значення числа піднесеного до степеня.

<u>Приклад 5.3.</u> Задано закон розподілу випадкової величини X :

X	-2	1	2
P	0,5	0,3	0,2

Знайти $M(X)$, $D(X)$.

Розв’язання.

$$M(X) = -2 \cdot 0,5 + 1 \cdot 0,3 + 2 \cdot 0,2 = -0,3;$$

$$D(X) = (-2+0,3)^2 \cdot 0,5 + (1+0,3)^2 \cdot 0,3 + (2+0,3)^2 \cdot 0,2 =$$

$$= 1,445 + 0,507 + 1,058 = 3,01.$$

B6 =СУММПРОИЗВ(C2:C4;B2:B4)-СТЕПЕНЬ(B5;2)

	A	B	C	D	E	F
1	X	P	X^2			
2	-2	0,5	4			
3	1	0,3	1			
4	2	0,2	4			
5	$M(X)$	-0,3				
6	$D(X)$	3,01				

ORIGIN ≡ 1

$X := (-2 \quad 1 \quad 2)$

$P := (0.5 \quad 0.3 \quad 0.2)$

$$M := \sum_{i=1}^{cols(X)} \left(X_{1,i} \cdot P_{1,i}\right) = -0.3$$

$$D := \left[\sum_{i=1}^{cols(X)} \left[\left(X_{1,i}\right)^2 \cdot P_{1,i}\right]\right] - M^2 = 3.01$$

Відповідь: $M(X) = -0,3$; $D(X) = 3,01$.

У деяких випадках зручніше обчислювати дисперсію за формулою:

$$D(X) = M(X^2) - M^2(X). \qquad (5.6)$$

Встановимо справедливість цієї рівності використовуючи властивості математичного сподівання:

$$D(X) = M\left[(X - M(X))^2\right] = M\left[X^2 - 2XM(X) + M^2(X)\right] =$$

$$= M(X^2) - 2M(X)M(X) + M^2(X) = M(X^2) - M^2(X).$$

<u>Приклад 5.4.</u> Обчислити за формулою (5.6) дисперсію для випадкової величини X з приклада 5.3.

<u>Розв'язання.</u> Запишемо закон розподілу для випадкової величини X^2 :

X^2	4	1	4
P	0,5	0,3	0,2

$M(X^2) = 4 \cdot 0,5 + 1 \cdot 0,3 + 4 \cdot 0,2 = 3,1.$

Отже, $D(X) = M(X^2) - M^2(X) = 3,1 - (0,3)^2 = 3,01.$

<u>Відповідь:</u> $D(X) = 3,01.$

Властивості дисперсії.

1. Дисперсія випадкової величини завжди невід'ємна, тобто $D(X) \geq 0.$ Ця властивість випливає з означення дисперсії.

2. Дисперсія сталої величини C дорівнює нулю, тобто $D(C) = 0.$

Дійсно, $D(C) = M\left[(C - M(C))^2\right] = M\left[(C - C)^2\right] = M(0) = 0.$

3. Сталий множник можна виносити за знак дисперсії, піднісши його при цьому до квадрата, тобто $D(CX) = C^2 D(X).$

Дійсно, $D(CX) = M\left[(CX - M(CX))^2\right] = M\left[(CX - CM(X))^2\right] =$

$$= M\left[C^2(X - M(X))^2\right] = C^2 M\left[(X - M(X))^2\right] = C^2 D(X).$$

4. Дисперсія алгебраїчної суми двох незалежних випадкових величин дорівнює сумі їх дисперсій, тобто $D(X \pm Y) = D(X) + D(Y).$

Застосуємо формулу 5.6:

$$D(X \pm Y) = M\left[(X \pm Y)^2\right] - M^2[(X \pm Y)] =$$

$$= M\left[(X^2 \pm 2XY + Y^2)\right] - (M(X) \pm M(Y))^2 =$$

$$= M(X^2) \pm 2M(XY) + M(Y^2) - M^2(X) \mp 2M(X)M(Y) - M^2(Y) =$$

$$= M(X^2) - M(X^2) + M(Y^2) - M(Y^2) = D(X) + D(Y).$$

При доведенні цієї формули ми скористалися властивістю 3^0 математичного сподівання добутку двох незалежних випадкових величин.

Наслідок. Якщо випадкову величину змінити на сталу величину, то дисперсія від цього не зміниться.

Дійсно, $D(X + C) = D(X) + D(C) = D(X) + 0 = D(X).$

Середнє квадратичне відхилення.

Для оцінки розсіювання можливих значень випадкової величини навколо її математичного сподівання застосовують числову характеристику, яка має ту ж розмірність, що і сама величина (на відміну від дисперсії, яка вимірюється у квадратних одиницях).

Означення. Середнім квадратичним відхиленням випадкової величини називається число, що дорівнює арифметичному значенню кореня квадратного з її дисперсії і позначається $\sigma(X)$, тобто $\sigma(X) = \sqrt{D(X)}.$

Середнє квадратичне відхилення випадкової величини як у *Excel*, так і у *MathCad* обчислюється безпосередньо за формулою, використовуючи:

- математичну функцію **КОРЕНЬ(число)**, яка повертає значення квадратного кореня числа.
- підпанель **Calculator** (Калькулятор) панелі **Math** (Математика).

<u>Приклад 5.5.</u> Прибуток акцій А та В залежить від стану економіки. Експерти вказали на 5 можливих станів економіки, а також ймовірності їх реалізації:

Економічний стан	ймовірність	норма прибутку акцій	
		А, %	В, %
1	2	3	4
значне піднесення	0,1	20	10
незначне піднесення	0,2	10	5
стагнація	0,3	2	2
незначна рецесія	0,3	-2	1
значна рецесія	0,1	-10	-5

Знайти очікувану норму прибутку кожної акції (математичне сподівання) і ризик (середнє квадратичне відхилення).

<u>Розв'язання.</u> Нехай випадкова величина X – норма прибутку акції А, Y – норма прибутку акції В.

Знаходимо $M(X)=20\cdot 0{,}1+10\cdot 0{,}2+2\cdot 0{,}3+(-2)\cdot 0{,}3+(-10)\cdot 0{,}1=3$,

$$M(Y)=10\cdot 0{,}1+5\cdot 0{,}2+2\cdot 0{,}3+1\cdot 0{,}3+(-5)\cdot 0{,}1=2{,}4.$$

Отже, очікуваний прибуток від акції А становить 3 %, а від акції В – 2,4 %.

Знайдемо ризики цих акцій

$$D(X)=400\cdot 0{,}1+100\cdot 0{,}2+4\cdot 0{,}3+4\cdot 0{,}3+100\cdot 0{,}1-3^2=63{,}4;$$

$$\sigma(X)=\sqrt{D(X)}=\sqrt{63{,}4}\approx 7{,}96;$$

$$D(X)=100\cdot 0{,}1+25\cdot 0{,}2+4\cdot 0{,}3+1\cdot 0{,}3+25\cdot 0{,}1-(2{,}4)^2=13{,}24;$$

$$\sigma(X)=\sqrt{D(X)}=\sqrt{13{,}24}\approx 3{,}6.$$

Таким чином хоча акція А і має більший очікуваний прибуток, але коливання її прибутковості теж більше, тому вона обтяжена більшим ризиком.

J16

	A	B	C	D	E	F	G	H	I
1	X	Y	P	X^2	Y^2				
2	20	10	0,1	400	100				
3	10	5	0,2	100	25				
4	2	2	0,3	4	4				
5	-2	1	0,3	4	1				
6	-10	-5	0,1	100	25				
7		$M(X)$	3	=СУММПРОИЗВ(A2:A6;C2:C6)					
8		$D(X)$	63,4	=СУММПРОИЗВ(D2:D6;C2:C6)-СТЕПЕНЬ(C7;2)					
9		$\sigma(X)$	7,9624	=КОРЕНЬ(C8)					
10		$M(Y)$	2,4	=СУММПРОИЗВ(B2:B6;C2:C6)					
11		$D(Y)$	13,24	=СУММПРОИЗВ(E2:E6;C2:C6)-СТЕПЕНЬ(C10;2)					
12		$\sigma(Y)$	3,6387	=КОРЕНЬ(C11)					

$$ORIGIN \equiv 1$$

$$X := (20 \;\; 10 \;\; 2 \;\; -2 \;\; -10)$$

$$Y := (10 \;\; 5 \;\; 2 \;\; 1 \;\; -5)$$

$$P := (0.1 \;\; 0.2 \;\; 0.3 \;\; 0.3 \;\; 0.1)$$

$$MX := \sum_{i=1}^{cols(X)} \left(X_{1,i} \cdot P_{1,i}\right) = 3 \qquad MY := \sum_{i=1}^{cols(Y)} \left(Y_{1,i} \cdot P_{1,i}\right) = 2.4$$

$$DX := \left[\sum_{i=1}^{cols(X)} \left[\left(X_{1,i}\right)^2 \cdot P_{1,i}\right]\right] - MX^2 = 63.4 \qquad DY := \left[\sum_{i=1}^{cols(Y)} \left[\left(Y_{1,i}\right)^2 \cdot P_{1,i}\right]\right] - MY^2 = 13.24$$

$$\sigma x := \sqrt{DX} = 7.962 \qquad \sigma y := \sqrt{DY} = 3.639$$

<u>Відповідь:</u> очікуваний прибуток від акції А становить 3 %, а від акції В – 2,4 %; хоча акція А і має більший очікуваний прибуток, але коливання її прибутковості теж більше, тому вона обтяжена більшим ризиком.

5.3. Числові характеристики дискретної випадкової величини основних законів розподілу

У цьому пункті будуть знайдені числові характеристики дискретних випадкових величин, які мають відомі закони розподілу і використовуються для побудови теоретико-ймовірнісних моделей реальних соціально-економічних явищ.

5.3.1. Біномний закон розподілу

<u>Теорема 5.1.</u> Математичне сподівання випадкової величини X, яка розподілена за біномним законом обчислюється за формулою

$$M(X) = np \tag{5.7}$$

а її дисперсія за формулою

$$D(X) = npq \tag{5.8}$$

<u>Доведення.</u> Випадкову величину X – числа появ події A в n незалежних випробуваннях схеми Бернуллі, можна подати у вигляді суми n незалежних випадкових величин $X_1 + X_2 + ... + X_n$, кожна з яких має один і той самий закон розподілу:

$X_k, k=\overline{1,n}$	0	1
P	q	p

Величина X_k набуває значення, що дорівнює числу появ події A в k-му випробуванні, тобто X_k може приймати значення або 1 – подія A відбулася, з ймовірністю p, або 0 – подія A не відбулася, з ймовірністю q.

Математичне сподівання і дисперсія випадкової величини X_k:

$$M(X_k) = 0 \cdot q + 1 \cdot p = p;$$

$$D(X_k) = M(X_k^2) - M^2(X_k) = p - p^2 = p(1-p) = pq.$$

Далі, використовуючи математичного сподівання і дисперсії, можна отримати:

$$M(X) = M(X_1 + X_2 + ... + X_n) = M(X_1) + M(X_2) + ... + M(X_n) =$$

$$= \underbrace{p + p + \ldots + p}_{n \text{ разів}} = np \ ;$$

$$D(X) = D(X_1 + X_2 + \ldots + X_n) = D(X_1) + D(X_2) + \ldots + D(X_n) =$$

$$= \underbrace{qp + qp + \ldots + qp}_{n \text{ разів}} = qnp \, . \bullet$$

5.3.2. Геометричний закон розподілу

<u>Теорема 5.2.</u> Математичне сподівання випадкової величини X, яка розподілена за геометричним законом, дорівнює $\frac{1}{p}$, а її дисперсія – $\frac{q}{p^2}$, тобто

$$M(X) = \frac{1}{p}; \tag{5.9}$$

$$D(X) = \frac{q}{p^2}. \tag{5.10}$$

<u>Доведення.</u> Якщо X має геометричний розподіл, то вона набуває значення $1, 2, \ldots$ і ймовірність того, що X прийме значення m обчислюється за формулою $P(X = m) = q^{m-1}p,$ отже,

$$M(X) = \sum_{m=1}^{\infty} mq^{m-1}p = p(1 + 2q + 3q^2 + \ldots + mq^{m-1} + \ldots) = p \cdot \frac{1}{(1-q)^2} = p \cdot \frac{1}{p^2} = \frac{1}{p}.$$

Тут було використано те, що $0 < q < 1$ і члени степеневого ряду $\sum_{m=1}^{\infty} q^{m-1}$ утворюють нескінченно спадну геометричну прогресію, тобто $\sum_{m=1}^{\infty} q^{m-1} = \frac{1}{1-q}$, а такий ряд можна почленно диференціювати і сума одержаного ряду дорівнюватиме похідній від суми початкового ряду, тобто

$$(1 + q + q^2 + \ldots + q^{m-1} + \ldots)' = \left(\frac{1}{1-q} \right)'$$

$$1 + 2q + 3q^2 + \ldots + (m-1)q^{m-2} + \ldots = \frac{1}{(1-q)^2} = \frac{1}{p^2}.$$

Для знаходження дисперсії $D(X)$, потрібно скористатись формулою $D(X) = M(X^2) - M^2(X)$ і тим, що

$$1^2+2^2q+3^2q^2+\ldots+n^2q^2+\ldots=\frac{1+q}{(1-q)^3}. \qquad (5.11)$$

Для того, щоб встановити, що рівність (5.11) виконується, потрібно ряд із (5.11) проінтегрувати почленно (це можливо, так як $0<q<1$):

$$\int_0^q\left(\sum_{n=1}^{\infty}n^2q^{n-1}\right)dq=\sum_{n=1}^{\infty}n^2\int_0^q q^{n-1}dq=\sum_{n=1}^{\infty}nq^n=q\sum_{n=1}^{\infty}nq^{n-1}=q\cdot\frac{1}{(1-q)^2}=\frac{q}{(1-q)^2}.$$

Рівність (5.11) отримується після диференціювання останньої рівності по q.

З рівності (5.11): $M(X^2)=p\sum_{n=1}^{\infty}n^2q^{n-1}=\frac{1+q}{p^2}.$

Отже,

$$D(X)=\frac{1+q}{p^2}-\frac{1}{p^2}=\frac{q}{p^2}. \bullet$$

5.3.3. Закон розподілу Пуассона

<u>Теорема 5.3.</u> Математичне сподівання і дисперсія випадкової величини, розподіленої за законом Пуассона, дорівнюють параметру λ цього закону, тобто

$$M(X)=\lambda; \qquad (5.12)$$

$$D(X)=\lambda. \qquad (5.13)$$

Доведення. Математичне сподівання:

$$M(X)=\sum_{k=1}^{\infty}k\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=1}^{\infty}\frac{\lambda^k e^{-\lambda}}{(k-1)!}=\lambda e^{-\lambda}\sum_{k=1}^{\infty}\frac{\lambda^{k-1}}{(k-1)!}=\lambda e^{-\lambda}\cdot e^{\lambda}=\lambda.$$

Для знаходження дисперсії, спочатку потрібно знайти

$$M(X^2)=\sum_{k=1}^{\infty}k^2\frac{\lambda^k}{k!}e^{-\lambda}=\sum_{k=1}^{\infty}\frac{k\lambda^k e^{-\lambda}}{(k-1)!}=e^{-\lambda}\sum_{k=1}^{\infty}\frac{((k-1)+1)\lambda^k}{(k-1)!}=$$

$$=e^{-\lambda}\left(\sum_{k=2}^{\infty}\frac{\lambda^k}{(k-2)!}+\sum_{k=1}^{\infty}\frac{\lambda^k}{(k-1)!}\right)=e^{-\lambda}\left(\lambda^2e^{\lambda}+\lambda e^{\lambda}\right)=\lambda^2+\lambda.$$

Отже,

$$D(X)=\lambda^2+\lambda-\lambda^2=\lambda. \bullet$$

5.4. Функція розподілу випадкової величини

Повне уявлення про випадкову величину може давати не тільки закон розподілу. Взагалі, коли випадкова величина набуває всі значення з певного числового проміжку, тобто є неперервною, її вже неможливо задати за допомогою закона розподілу, так як, по-перше, вона приймає незліченну множину значень, по-друге, як видно буде далі, ймовірність кожного окремого значення неперервної випадкової величини дорівнює нулю. Отже, можливий інший підхід для задання випадкової величини: розглянути не ймовірність події $X = x$ для різних x (як це робилось у законі розподілу), а ймовірність події $X < x$, де x – змінна. Ймовірність $P(X < x)$, очевидно, залежить від x, тобто є функцією від x.

Означення. *Функцією розподілу* $F(x)$ випадкової величини X називається ймовірність того, що X набуває значення, меншого значення x, де x – довільне дійсне число:

$$F(x) = P(X < x).$$

Функцію $F(x)$ інколи називають *інтегральною функцією розподілу* або *інтегральним законом розподілу*.

Властивості функції розподілу.

1. Функція розподілу випадкової величини набуває свої значення з проміжка $[0;1]$, тобто

$$0 \le F(x) \le 1.$$

Це твердження випливає з того, що функція розподілу – це ймовірність.

2. Функція розподілу $F(x)$ – неспадна, тобто $F(x_2) \ge F(x_1),$ якщо $x_2 \ge x_1$.

Доведення. Нехай $x_2 > x_1$. Потрібно розглянути дві несумісні події $A = \{X < x_1\}$ і $B = \{x_1 \le X < x_2\}$. Тоді $A + B = \{X < x_2\}$. За теоремою додавання несумісних подій:

$$P(A+B) = P(A) + P(B) \text{ або } P(X < x_2) = P(X < x_1) + P(x_1 < X < x_2).$$

Отже,

$$F(x_2)=F(x_1)+P(x_1\le X<x_2). \qquad (5.14)$$

Оскільки $P(x_1\le X<x_2)\ge 0$, то $F(x_2)\ge F(x_1)$, тобто $F(x)$ – неспадна.

3. Ймовірність того, що випадкова величина набуде значення з проміжку $[x_1, x_2)$ дорівнює приросту на цьому проміжку:

$$P(x_1\le X<x_2)=F(x_2)-F(x_1).$$

Ця властивість одразу випливає з рівності (5.14).

4. Якщо $x\to+\infty$, то $F(x)\to 1$ і якщо $x\to-\infty$, то $F(x)\to 0$.

Доведення. $\lim\limits_{x\to+\infty}F(x)=F(+\infty)=P(X<+\infty)=1$, як ймовірність достовірної події. $\lim\limits_{x\to-\infty}F(x)=F(-\infty)=P(X<-\infty)=0$, як ймовірність неможливої події. •

Для знаходження та побудови функції розподілу:

 створюється модуль, у якому використовуються функції та оператори підпанелі **Programming** (Программирование) панелі **Math** (Математика):

Add Line – викликає шаблон вертикальної лінії для вводу на місці міток потрібних операторів. Вертикальна лінія означає, що оператори, які до неї відносяться, будуть утворювати один блок;

$\leftarrow$ – символ локального присвоювання (у тілі модуля);

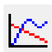
for – викликає шаблон оператора циклу з фіксованим числом повторень;

while – викликає шаблон оператора циклу з умовою;

break – викликає шаблон оператора переривання;

if – викликає шаблон умовного оператора;

otherwise – викликає шаблон оператора іншого вибору (найчастіше зустрічається із оператором *if*).

Для побудови графіка використовується оператор  підпанелі **Graph** (Графики) панелі **Math** (Математика).

<u>Приклад 5.6.</u> Задано закон розподілу дискретної випадкової величини X :

X	1	3	5
P	0,2	0,3	0,5

Знайти функцію розподілу і побудувати її графік.

<u>Розв'язання.</u>

1. Якщо $x \le 1$, то $F(x) = P(X < 1) = 0$ (неможлива подія).

2. Якщо $1 < x \le 3$, то $F(x) = P(X < 3) = P(X = 1) = 0,2$.

3. Якщо $3 < x \le 5$, то $F(x) = P(X < 5) = P(X = 1) + P(X = 3) = 0,5$.

4. Якщо $x > 5$, то $F(x) = P(X = 1) + P(X = 3) + P(X = 5) = 1$ (достовірна подія).

Отже,

$$F(x) = \begin{cases} 0, & \text{при } x \le 1; \\ 0,2, & \text{при } 1 < x \le 3; \\ 0,5, & \text{при } 3 < x \le 5; \\ 1, & \text{при } x > 5. \end{cases}$$

Графік цієї функції зображено на рис. 5.1:

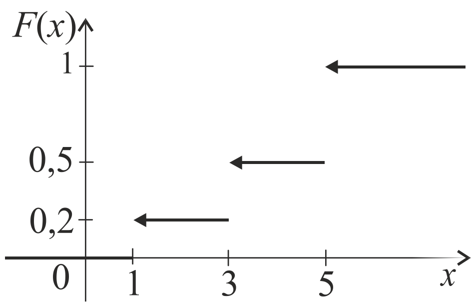
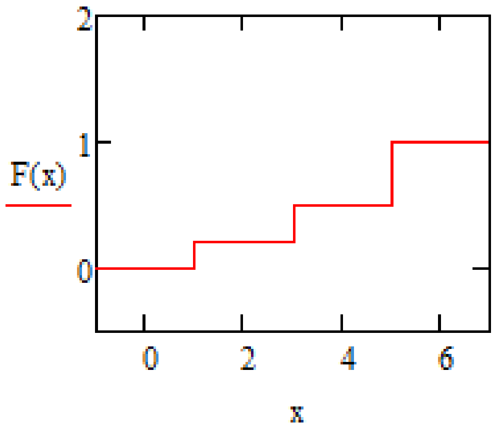


Рис. 5.1

ORIGIN ≡ 1

X := (1 3 5)   P := (0.2 0.3 0.5)

```
F(x) := | p ← 0
        | i ← 1
        | while x > X_{1,i}
        |     | p ← p + P_{1,i}
        |     | i ← i + 1
        |     | break if i > cols(X)
        | p
```



Практичне заняття 5

5.1. Знайти математичне сподівання випадкової величини $Z = 6X - 3Y$, якщо $M(X) = 2$ і $M(Y) = 5$.

5.2. Знайти математичне сподівання випадкової величини $Z = 2X - 7Y$, якщо $M(X) = 1$ і $M(Y) = 2$.

5.3. Задано закон розподілу дискретної випадкова величина X :

X	0	1	2
P	0,3	0,5	0,2

Знайти $M(X)$, $D(X)$, $\sigma(X)$.

5.4. Задано закон розподілу дискретної випадкова величина X :

X	-1	0	2	4
P	0,2	0,1	p_3	0,3

Знайти p_3, $M(X)$, $D(X)$, $\sigma(X)$.

5.5. Задано закон розподілу дискретної випадкова величина X :

X	-1	2	7
P	p_1	0,4	0,3

Знайти p_1, $M(X)$, $D(X)$, $\sigma(X)$.

5.6. Задано закон розподілу дискретної випадкова величина X :

X	-2	3	4
P	0,7	p_2	0,2

Знайти p_2, $M(X)$, $D(X)$, $\sigma(X)$.

5.7. Відомо, що дискретна випадкова величина X набуває трьох значень, причому $x_2 = 1$ з ймовірністю $p_2 = 0,5$ та $x_3 = 2$ з ймовірністю $p_3 = 0,2$. Знайти x_1, p_1, $D(X)$, якщо відомо, що $M(X) = 0$.

5.8. Відомо, що дискретна випадкова величина X набуває трьох значень, причому $x_1 = 1$ з ймовірністю $p_1 = 0,5$ та $x_3 = 2$ з ймовірністю $p_3 = 0,2$. Знайти x_2, p_2, $D(X)$, якщо відомо, що $M(X) = 0,8$.

5.9. Відомо, що дискретна випадкова величина X набуває трьох значень: $x_1 = 1$, $x_2 = 3$, $x_3 = 5$. Знайти p_1, p_2, p_3, якщо відомо, що $M(X) = 1,8$; $M(X^2) = 5$.

5.10. Відомо, що дискретна випадкова величина X набуває трьох значень: $x_1 = -1$, $x_2 = 0$, $x_3 = 3$. Знайти p_1, p_2, p_3, якщо відомо, що $M(X) = 0$; $M(X^2) = 1,2$.

5.11. Підприємство, на якому працює автоматизована лінія по розливу соків, періодично організовує перевірки. Остання перевірка ваги пакетів соку, розлитого по 1000 мл, дала наступні результати:

Вага пакету, мл	998	999	1000	1001	1002
Кількість пакетів	5	9	50	11	8

Обчислити середню вагу пакетів соку, розлитого автоматизованою лінією.

5.12. Підприємство вирішило отримати прибуток, вклавши гроші у три інноваційні проекти, розвиток яких залежить як від стану економіки, так і від впливу інших різних чинників. Експерти вказали на 2 можливих стани розвитку ситуацій, а також ймовірності їх реалізації.

Стан розвитку ситуації	ймовірність	норма прибутку від інноваційного проекту		
		I, %	II, %	III, %
прогнозований	0,7	30	20	50
катастрофічний	0,3	-10	-5	-20

Знайти очікувану норму прибутку кожного інноваційного проекту (математичне сподівання) і ризик (середнє квадратичне відхилення).

5.13. Прибуток підприємства, яке підготувало для реалізації два види нових миючих засобів А та В, залежить від попиту на них. Маркетологи прогнозують виникнення 4 видів попиту при реалізації нових миючих засобів, а також ймовірності, з якими вони можуть настати:

Вид попиту	ймовірність	норма прибутку від реалізації нового товару	
		А, %	В, %
1	**2**	**3**	**4**
повноцінний	0,5	15	12
потенційний	0,3	10	8

1	2	3	4
прихований	0,1	4	3
негативний	0,1	-3	-1

Знайти очікувану норму прибутку кожного миючого засобу (математичне сподівання) і ризик (середнє квадратичне відхилення).

5.14. Ймовірність отримати прибуток від акції становить 0,3. Визначити математичне сподівання та дисперсію числа прибуткових акцій із 45 придбаних.

5.15. Бухгалтер здає звіт, який містить 100 сторінок, у податкову інспекцію. Ймовірність того, що на сторінці звіту є неточності дорівнює 0,2. Визначити математичне сподівання та дисперсію числа сторінок звіту, які не містять неточностей.

5.16. Митний контроль сповіщає, що 15 % усіх осіб, що повертаються з-за кордону, не декларують товар, який підлягає оподаткуванню. Відомо, що за останню зміну повернулось 135 осіб. Знайти математичне сподівання та дисперсію числа осіб, які не задекларували ввезений товар, підлягаючий оподаткуванню.

5.16. Проводяться багаторазові випробування експериментального апарату на надійність до тих пір, поки він не відмовить у роботі. Ймовірність відмови апарату в кожному випробуванні дорівнює 0,15. Знайти числові характеристики дискретної випадкової величини X – числа випробувань, які потрібно провести для перевірки апарату на надійність.

5.17. Викладач задає питання студенту до тих пір, поки він не дасть повну відповідь. Відомо, що ймовірність того, що студент дасть повну відповідь на питання становить 0,3. Знайти математичне сподівання та дисперсію числа заданих викладачем питань.

5.18. Контролер звіряє показання лічильників власників квартир до тих пір, поки не виявить помилку. Ймовірність виявлення помилки контролером дорівнює 0,1. Знайти математичне сподівання та дисперсію числа перевірених показань лічильників контролером.

5.19. Підприємство відправило замовнику 3000 якісних товарів. Ймовірність пошкодження товару в дорозі дорівнює 0,001. Знайти математичне сподівання та дисперсію числа пошкоджених товарів.

5.20. Відомий географ написав мемуари, які відправив у видавництво. Ймовірність того, що оператор при наборі мемуарів допустить помилку у слові дорівнює 0,005. Мемуари містять 100000 слів. Знайти середнє число слів, які будуть містити помилку після набору оператором.

5.21. Відомо, що в одному досить освіченому містечку 99,999 % жителів прочитали хоча б одну книгу. Для соціологічного дослідження відібрали 1000 жителів. Знайти числові характеристики дискретної випадкової величини X – числа жителів, які не прочитали жодної книги.

5.22. Задано закон розподілу дискретної випадкової величини X :

X	-1	0	2
P	0,1	0,6	0,3

Знайти функцію розподілу і побудувати її графік.

5.23. Задано закон розподілу дискретної випадкової величини X :

X	-1	1	4
P	0,2	0,1	0,7

Знайти функцію розподілу і побудувати її графік.

5.24. Задано закон розподілу дискретної випадкової величини X :

X	1	2	4
P	0,1	0,4	0,5

Знайти функцію розподілу і побудувати її графік.

5.25. Задано закон розподілу дискретної випадкової величини X :

X	1	2	4	7
P	0,2	0,1	0,4	0,3

Знайти функцію розподілу і побудувати її графік.

5.26. Задано закон розподілу дискретної випадкової величини X :

X	-2	0	3	9
P	0,5	0,2	0,1	0,2

Знайти функцію розподілу і побудувати її графік.

Лекція 6. Неперервні випадкові величини

6.1. Означення неперервної випадкової величини

6.2. Щільність розподілу ймовірностей неперервної випадкової величини

6.3. Числові характеристики неперервної випадкової величини

6.1. Означення неперервної випадкової величини

У попередній лекції було розглянуто випадкові величини, які можуть набувати скінчене або злічене число значень. Якщо значеннями випадкової величини є деякий проміжок (скінчений або нескінчений) числової осі, тоді таку випадкову величину називають неперервною.

Приклади неперервної випадкової величини:

- дальність польоту снаряду;
- витрати електроенергії на підприємстві за місяць.

Далі подано більш строге означення неперервної випадкової величини.

<u>*Означення.*</u> *Випадкова величина* X називається *неперервною*, якщо її функція розподілу $F(x)$ є неперервною і диференційованою всюди, крім, можливо, окремих точок.

Для неперервної випадкової величини виконується наступне твердження.

<u>***Теорема.***</u> Ймовірність того, що неперервна випадкова величина X приймає одне певне значення дорівнює нулю.

<u>Доведення.</u> Нехай $a = x$ і $b = x + \Delta x$, тоді $P(x \le X < x + \Delta x) = F(x + \Delta x) - F(x)$. Якщо $\Delta x \to 0$, то $F(x + \Delta x) - F(x) \to 0$. Отже, $P(x) = 0$.•

<u>***Наслідок.***</u> Якщо X неперервна випадкова величина, то

$$P(a \le X \le b) = P(a \le X < b) = P(a < X \le b) = P(a < X < b).$$

Для побудови функції розподілу:

створюється модуль, описаний у лекції 5 (с. 72 – 73).

<u>Приклад 6.1.</u> Неперервна випадкова величина X задана функцією розподілу:

$$F(x)=\begin{cases}0, & \text{при } x\le 0;\\ \dfrac{x^2}{4}, & \text{при } 0<x\le 2;\\ 1, & \text{при } x>2.\end{cases}$$

Побудувати графік функції $F(x)$ і знайти ймовірність того, що випадкова величина X набуде значення з інтервалу $(0;1)$.

<u>Розв'язання.</u> Графік функції $F(x)$ зображено на рис. 6.1.

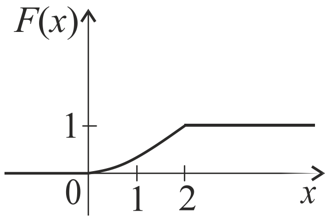
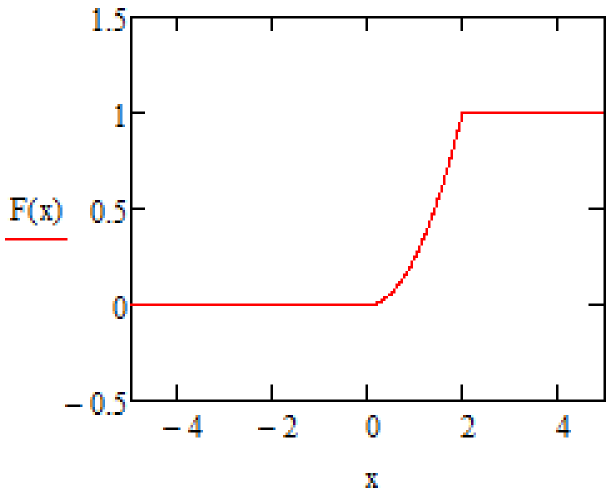


Рис. 6.1

Ймовірність того, що випадкова величина X набуде значення з інтервалу $(0;1)$:

$$P(0<X<1)=F(1)-F(0)=\frac{1}{4}-0=\frac{1}{4}.$$

$$F(x) := \begin{array}{|l} 0 \ \text{if} \ x\le 0\\ \dfrac{x^2}{4} \ \text{if} \ 0<x\le 2\\ 1 \ \text{if} \ x>2\end{array}$$



$$P := F(1) - F(0) = 0.25$$

Відповідь: $P(0 < X < 1) = \dfrac{1}{4}$.

6.2. Щільність розподілу ймовірностей неперервної випадкової величини

Задання випадкової величини за допомогою функції розподілу не є єдиним. Далі буде розглянуте поняття щільності розподілу ймовірностей неперервної випадкової величини.

Ймовірність того, що неперервна випадкова величина набуде значення з відрізку $[x; x+\Delta x]$ дорівнює $P(x \le X < x+\Delta x) = F(x+\Delta x) - F(x)$, тобто дорівнює приросту функції $F(x)$ на цьому відрізку. Отже,

$$\lim_{\Delta x \to 0} \frac{P(x \le X \le x+\Delta x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x+\Delta x) - F(x)}{\Delta x} = F'(x).$$

Означення. *Щільністю розподілу ймовірностей* $f(x)$ (або *диференціальною функцією розподілу*) неперервної випадкової величини X називається похідна її функції розподілу:

$$f(x) = F'(x). \tag{6.1}$$

Для щільності розподілу:

при знаходженні похідної використовується оператор підпанелі **Calculus (Вычисления)** панелі **Math (Математика)**.

<u>**Приклад 6.2.**</u> Знайти щільність розподілу ймовірностей випадкової величини X, заданої функцією розподілу з <u>прикладу 6.1,</u> і побудувати її графік.

<u>Розв'язання.</u> Щільність розподілу ймовірностей випадкової величини X знаходиться за формулою (6.1):

$$f(x) = F'(x) = \begin{cases} 0, & \text{при } x \le 0; \\ \dfrac{x}{2}, & \text{при } 0 < x \le 2; \\ 0, & \text{при } x > 2. \end{cases}$$

Графік щільності розподілу ймовірностей випадкової величини X зображено на рис. 6.2.

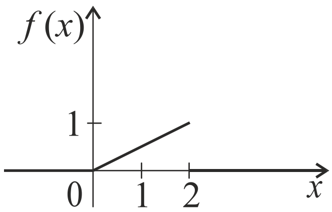
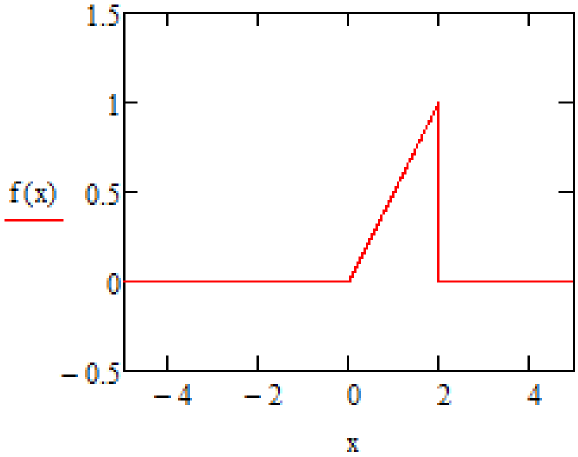


Рис. 6.2

$$F(x) := \begin{vmatrix} 0 \ \text{if} \ x \le 0 \\ \dfrac{x^2}{4} \ \text{if} \ 0 < x \le 2 \\ 1 \ \text{if} \ x > 2 \end{vmatrix}$$

$$f(x) := \frac{d}{dx}F(x)$$



Відповідь: $f(x)=\begin{cases}0, & \text{при } x\le 0;\\ \dfrac{x}{2}, & \text{при } 0<x\le 2;\\ 0, & \text{при } x>2.\end{cases}$

Властивості щільності розподілу ймовірностей неперервної випадкової величини.

1^0. Щільність розподілу ймовірностей випадкової величини є невід'ємною функцією, тобто $f(x)\ge 0$.

Дійсно, це випливає з того, що функція $F(x)$ є неспадною, тому $F'(x)\ge 0$, тобто $f(x)\ge 0$.

2^0. Ймовірність того, що неперервна випадкова величина набуде значення з інтервалу $(a;b)$ дорівнює визначеному інтегралу від щільності розподілу ймовірностей в межах від a до b:

$$P(a<X<b)=\int_a^b f(x)dx. \tag{6.2}$$

Дійсно, з властивості функції розподілу $F(x)$: $P(a<X<b)=F(b)-F(a)$, а з означення щільності розподілу ймовірностей: $\int_a^b f(x)dx = F(x)\Big|_a^b = F(b)-F(a).$

Отже, $P(a<X<b)=\int_a^b f(x)dx.$

3^0. Функція розподілу неперервної випадкової величини виражається через щільність розподілу ймовірностей $f(x)$ за допомогою рівності

$$F(x) = \int_{-\infty}^{x} f(t)dt. \tag{6.3}$$

Дійсно, з означення $F(x) = P(X < x)$, тобто $-\infty < X < x$, тому згідно з формулою (6.2):

$$F(x) = P(-\infty < X < x) = \int_{-\infty}^{x} f(t)dt.$$

4^0. Невласний інтеграл з нескінченими межами інтегрування від щільності розподілу ймовірностей дорівнює одиниці:

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

З формули (6.3): $F(x) = \int_{-\infty}^{x} f(t)dt,$ якщо $x \to +\infty$, то $F(+\infty) = 1$.

6.3. Числові характеристики неперервної випадкової величини

У випадку неперервних випадкових величин, як і у випадку дискретних, використовуються числові характеристики: математичне сподівання $(M(X))$, дисперсія $(D(X))$, середнє квадратичне відхилення $(\sigma(X))$. Для отримання відповідних формул для $M(X)$ і $D(X)$ достатньо в формулах (5.1) і (5.5) для дискретної випадкової величини X замінити знак суми $\sum_{i=1}^{n}$ за усіма її значеннями знаком інтегралу з нескінченими межами інтегрування, дискретний аргумент x_i на неперервний x, а ймовірність p_i – елементом ймовірності $f(x)dx$.

У результаті можна одержати наступні формули для математичного сподівання і дисперсії неперервної випадкової величини X :

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx; \tag{6.4}$$

$$D(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 f(x) dx. \qquad (6.5)$$

Зауваження 6.1. Невласний інтеграл у формулі (6.3) вважається абсолютно збіжним, а у формулі (6.5) збіжним.

Зауваження 6.2. Якщо неперервна випадкова величина набуває усі свої значення з відрізку $[a;b]$, то її математичне сподівання і дисперсія знаходяться відповідно за формулами:

$$M(X) = \int_a^b xf(x)dx; \qquad (6.6)$$

$$D(X) = \int_a^b (x - M(X))^2 f(x) dx. \qquad (6.7)$$

Як і у випадку дискретних випадкових величин, дисперсію неперервної випадкової величини X доцільно знаходити за формулами:

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x) dx - M^2(X); \qquad (6.5^*)$$

$$D(X) = \int_a^b x^2 f(x) dx - M^2(X). \qquad (6.7^*)$$

Середнє квадратичне відхилення знаходиться за формулою: $\sigma(X) = \sqrt{D(X)}$.

Зауваження 6.3. Усі числові характеристики неперервної випадкової величини мають властивості аналогічні властивостям відповідних числових характеристик дискретних випадкових величин.

Для застосування властивостей щільності розподілу:

при знаходженні визначеного інтегралу використовуються оператори та підпанелі **Calculus (Вычисления)** панелі **Math (Математика)**.

Приклад 6.3. Функція $f(x)$ задана у вигляді

$$f(x)=\begin{cases}0, & \text{при } x\le 0;\\ Cx, & \text{при } 0<x\le 1;\\ 0, & \text{при } x>1.\end{cases}$$

Знайти: 1) значення параметра C, при якому задана функція буде щільністю розподілу ймовірностей деякої випадкової величини X; 2) $M(X)$ і $D(X)$.

<u>Розв’язання.</u> 1) З властивості 4^0:

$$\int_{-\infty}^{+\infty} f(x)dx=\int_0^1 Cxdx = C\cdot\left.\frac{x^2}{2}\right|_0^1 = C\cdot\frac{1}{2}=1,\ \text{звідки } C=2.$$

Тоді $f(x)=\begin{cases}0, & \text{при } x\le 0;\\ 2x, & \text{при } 0<x\le 1;\\ 0, & \text{при } x>1.\end{cases}$

2) Математичне сподівання та дисперсію можна знайти за формулами (6.6) та (6.7*) відповідно:

$$M(X)=\int_0^1 x\cdot 2xdx = 2\int_0^1 x^2dx = 2\cdot\left.\frac{x^3}{3}\right|_0^1=\frac{2}{3};$$

$$D(X)=\int_0^1 x^2\cdot 2xdx-\left(\frac{2}{3}\right)^2=\left.\frac{x^4}{4}\right|_0^1-\frac{4}{9}=\frac{1}{18}.$$

$$f(x,C) := \begin{array}{|l} C\cdot x \ \ \text{if } \ 0<x\le 1\\ 0 \ \ \text{otherwise}\end{array}$$

$$C := \frac{1}{\displaystyle\int_0^1 f(x,1)\,dx} \qquad C = 2$$

$$M := \int_{-\infty}^{\infty} x\cdot f(x,C)\,dx = 0.667$$

$$D := \int_{-\infty}^{\infty} x^2 \cdot f(x, C)\, dx - M^2 = 0.056$$

Відповідь: 1) $C = 2$; 2) $M(X) = \frac{2}{3}$, $D(X) = \frac{1}{18}$.

Практичне заняття 6

6.1. Неперервна випадкова величина X задана функцією розподілу:

$$F(x) = \begin{cases} 0, & \text{при } x \le 0; \\ \frac{1}{16}(8x - x^2), & \text{при } 0 < x \le 4; \\ 1, & \text{при } x > 4. \end{cases}$$

Знайти ймовірність того, що у результаті випробування неперервна випадкова величина X набуде значення:

1) меншого -1; 2) меншого 3; 3) не меншого 3; 4) не меншого 5; 5) від -1 до 3.

6.2. Неперервна випадкова величина X задана функцією розподілу:

$$F(x) = \begin{cases} 0, & \text{при } x \le -1; \\ -x^2 + 1, & \text{при } -1 < x \le 0; \\ 1, & \text{при } x > 0. \end{cases}$$

Знайти ймовірність того, що у результаті випробування неперервна випадкова величина X набуде значення:

1) меншого -3; 2) меншого 0,5; 3) не меншого 0,5; 4) не меншого 3; 5) від -2 до 0.

6.3. Неперервна випадкова величина X задана функцією розподілу:

$$F(x)=\begin{cases}0, & \text{при } x\le 1;\\ x-1, & \text{при } 1<x\le 2;\\ 1, & \text{при } x>2.\end{cases}$$

1) Знайти щільність розподілу $f(x)$, побудувати графіки функцій $F(x)$ та $f(x)$; 2) знайти ймовірність того, що неперервна випадкова величина X набуде значення з інтервалу $(1;1{,}5)$.

6.4. Неперервна випадкова величина X задана функцією розподілу:

$$F(x)=\begin{cases}0, & \text{при } x\le -1;\\ \frac{1}{4}(x+1), & \text{при } -1<x\le 3;\\ 1, & \text{при } x>3.\end{cases}$$

1) Знайти щільність розподілу $f(x)$, побудувати графіки функцій $F(x)$ та $f(x)$; 2) знайти ймовірність того, що неперервна випадкова величина X набуде значення з інтервалу $(0;2)$.

6.5. Неперервна випадкова величина X задана функцією розподілу:

$$F(x)=\begin{cases}0, & \text{при } x\le 0;\\ \frac{x^2}{9}, & \text{при } 0<x\le 3;\\ 1, & \text{при } x>3.\end{cases}$$

1) Знайти щільність розподілу $f(x)$, побудувати графіки функцій $F(x)$ та $f(x)$; 2) знайти ймовірність того, що неперервна випадкова величина X набуде значення з інтервалу $(1;2)$.

6.6. Неперервна випадкова величина X задана функцією розподілу:

$$F(x)=\begin{cases}0, & \text{при } x<0;\\ -x^2+2x, & \text{при } 0\le x<1;\\ 1, & \text{при } x\ge 1.\end{cases}$$

1) Знайти щільність розподілу $f(x)$, побудувати графіки функцій $F(x)$ та $f(x)$; 2) знайти ймовірність того, що неперервна випадкова величина X набуде значення з інтервалу $(0;0,5)$.

6.7. Неперервна випадкова величина X задана функцією розподілу:

$$F(x)=\begin{cases}0, & \text{при } x<-0,5;\\ (2x+1)^2, & \text{при } -0,5\le x<0;\\ 1, & \text{при } x\ge 0.\end{cases}$$

1) Знайти щільність розподілу $f(x)$, побудувати графіки функцій $F(x)$ та $f(x)$; 2) знайти ймовірність того, що неперервна випадкова величина X набуде значення з інтервалу $(-0,5;0,5)$; 3) знайти математичне сподівання та дисперсію неперервної випадкової величини X.

6.8. Неперервна випадкова величина X задана функцією розподілу:

$$F(x)=\begin{cases}0, & \text{при } x\le 2;\\ (2x-4)^2, & \text{при } 2<x\le {}^5\!/_2;\\ 1, & \text{при } x>{}^5\!/_2.\end{cases}$$

1) Знайти щільність розподілу $f(x)$, побудувати графіки функцій $F(x)$ та $f(x)$; 2) знайти ймовірність того, що неперервна випадкова величина X набуде значення з інтервалу $(2;2,2)$; 3) знайти математичне сподівання та дисперсію неперервної випадкової величини X.

6.9. Функція $f(x)$ задана у вигляді

$$f(x)=\begin{cases}0, & \text{при } x<-2;\\ C(x+1), & \text{при } -2\le x<-1;\\ 0, & \text{при } x\ge -1.\end{cases}$$

Знайти: 1) значення параметра С, при якому задана функція буде щільністю розподілу ймовірностей деякої випадкової величини X; 2) математичне сподівання та дисперсію випадкової величини X.

6.10. Функція $f(x)$ задана у вигляді

$$f(x)=\begin{cases}0, & \text{при } x<2;\\ C(x-3), & \text{при } 2\le x<3;\\ 0, & \text{при } x\ge 3.\end{cases}$$

Знайти: 1) значення параметра С, при якому задана функція буде щільністю розподілу ймовірностей деякої випадкової величини X; 2) математичне сподівання та дисперсію випадкової величини X.

6.11. Функція $f(x)$ задана у вигляді

$$f(x)=\begin{cases}0, & \text{при } x<0;\\ C(x-1), & \text{при } 0\le x<1;\\ 0, & \text{при } x\ge 1.\end{cases}$$

1) Знайти значення параметра С, при якому задана функція буде щільністю розподілу ймовірностей деякої випадкової величини X; 2) знайти функцію розподілу випадкової величини X; 3) побудувати графіки функцій $F(x)$ і $f(x)$; 4) знайти математичне сподівання та дисперсію випадкової величини X.

6.12. Функція $f(x)$ задана у вигляді

$$f(x)=\begin{cases}0, & \text{при } x<6;\\ C(x-7), & \text{при } 6\le x<7;\\ 0, & \text{при } x\ge 7.\end{cases}$$

1) Знайти значення параметра С, при якому задана функція буде щільністю розподілу ймовірностей деякої випадкової величини X; 2) знайти функцію розподілу випадкової величини X; 3) побудувати графіки функцій $F(x)$ і $f(x)$; 4) знайти математичне сподівання та дисперсію випадкової величини X.

6.13. Неперервна випадкова величина X задана щільністю розподілу:

$$f(x)=\begin{cases}0, & \text{при } x<0;\\ 2x, & \text{при } 0\le x<1;\\ 0, & \text{при } x\ge 1.\end{cases}$$

1) Знайти функцію розподілу $F(x)$ та побудувати графіки функцій $F(x)$ і $f(x)$; 2) знайти математичне сподівання та дисперсію неперервної випадкової величини X.

6.14. Неперервна випадкова величина X задана щільністю розподілу:

$$f(x)=\begin{cases}0, & \text{при } x\le 2;\\ 2x-4, & \text{при } 2<x\le 3;\\ 0, & \text{при } x>3.\end{cases}$$

1) Знайти функцію розподілу $F(x)$ та побудувати графіки функцій $F(x)$ і $f(x)$; 2) знайти математичне сподівання та дисперсію неперервної випадкової величини X.

6.15. Неперервна випадкова величина X задана щільністю розподілу:

$$f(x)=\begin{cases}0, & \text{при } x \le 1;\\ x-\dfrac{1}{2}, & \text{при } 1 < x \le 2;\\ 0, & \text{при } x > 2.\end{cases}$$

1) Знайти функцію розподілу $F(x)$ та побудувати графіки функцій $F(x)$ і $f(x)$; 2) знайти математичне сподівання та дисперсію неперервної випадкової величини X; 3) знайти ймовірність того, що неперервна випадкова величина X набуде значення з інтервалу $(1;3)$.

6.16. Неперервна випадкова величина X задана щільністю розподілу:

$$f(x)=\begin{cases}0, & \text{при } x \le 0;\\ \dfrac{3}{4}(x^2+1), & \text{при } 0 < x \le 1;\\ 0, & \text{при } x > 1.\end{cases}$$

1) Знайти функцію розподілу $F(x)$ та побудувати графіки функцій $F(x)$ і $f(x)$; 2) знайти математичне сподівання та дисперсію неперервної випадкової величини X; 3) знайти ймовірність того, що неперервна випадкова величина X набуде значення з інтервалу $(0;0,5)$.

Лекція 7. Закони розподілу неперервних випадкових величин

7.1. Рівномірний закон розподілу

7.2. Показниковий закон розподілу

7.3. Нормальний закон розподілу

7.4. Ймовірність попадання значення нормально розподіленої випадкової величини в інтервал. «Правило трьох сигм»

7.5. Закони розподілів χ^2, Стьюдента та Фішера-Снедекора

7.1. Рівномірний закон розподілу

<u>*Означення.*</u> Неперервна випадкова величина X має *рівномірний закон розподілу* на відрізку $[a;b]$, якщо її щільність розподілу ймовірностей $f(x)$ стала на цьому відрізку і має вигляд

$$f(x)=\begin{cases}0, & \text{при } x\le a;\\ \dfrac{1}{b-a}, & \text{при } a<x\le b;\\ 0, & \text{при } x>b.\end{cases}$$

<u>***Теорема.***</u> Функція розподілу випадкової величини X з рівномірним законом розподілу має вигляд

$$F(x)=\begin{cases}0, & \text{при } x\le a;\\ \dfrac{x-a}{b-a}, & \text{при } a<x\le b;\\ 1, & \text{при } x>b.\end{cases}$$

а її математичне сподівання і дисперсія відповідно дорівнюють

$$M(X)=\frac{a+b}{2};\ D(X)=\frac{(b-a)^2}{12}. \tag{7.1}$$

<u>Доведення.</u>

Дійсно, при $x\le a$, $F(x)=0$;

при $a<x\le b$, $F(x)=\int\limits_a^x\frac{dt}{b-a}=\left.\frac{t}{b-a}\right|_a^x=\frac{x-a}{b-a}$;

$$\text{при } x > b,\ F(x) = \int_a^b \frac{dx}{b-a} = \frac{x}{b-a}\bigg|_a^b = 1;$$

математичне сподівання $M(X) = \int_a^b \frac{x}{b-a}dx = \frac{1}{b-a} \cdot \frac{x^2}{2}\bigg|_a^b = \frac{a+b}{2}$;

дисперсія $D(X) = \int_a^b \frac{x^2}{b-a}dx - \left(\frac{a+b}{2}\right)^2 = \frac{1}{b-a} \cdot \frac{x^3}{3}\bigg|_a^b - \frac{(a+b)^2}{4} = \frac{b^3-a^3}{3(b-a)} -$

$$-\frac{a^2+2ab+b^2}{4} = \frac{b^2+ab+a^2}{3} - \frac{a^2+2ab+b^2}{4} = \frac{a^2-2ab+b^2}{12} = \frac{(a-b)^2}{12}. \bullet$$

Графіки функцій $f(x)$ і $F(x)$ зображені на рис. 7.1 (а, б відповідно):

а) б)

Рис. 7.1

Для рівномірного закону розподілу:

при знаходженні щільності розподілу використовується функція **dunif(*x*,*a*,*b*)**, функції розподілу – **punif(*x*,*a*,*b*)**, де ***x*** – значення випадкової величини на відрізку $[a;b]$.

<u>**Приклад 7.1.**</u> Поїзди метрополітену ходять регулярно із проміжком 3 хв. Яка ймовірність того, що пасажиру прийдеться чекати поїзд не більше 1 хв. Знайти математичне сподівання і середнє квадратичне відхилення випадкової величини X – часу очікування поїзда.

Розв’язання. Випадкова величина X – час очікування поїзда на часовому відрізку $[0;3]$ має рівномірний розподіл зі щільність розподілу ймовірностей $f(x)=\frac{1}{3}$. Тому $P(X\le 1)=\int_0^1\frac{1}{3}dx=\frac{1}{3}x\Big|_0^1=\frac{1}{3}$.

З формул (7.1):

$$M(X)=\frac{0+3}{2}=1{,}5;\ D(X)=\frac{(3-0)^2}{12}=\frac{3}{4};\ \sigma(X)=\frac{\sqrt{3}}{2}.$$

$$a := 0 \quad b := 3$$

$$M := \int_{-\infty}^{\infty} x\cdot \mathrm{dunif}(x,a,b)\,dx = 1.5$$

$$D := \int_{-\infty}^{\infty} x^2\cdot \mathrm{dunif}(x,a,b)\,dx - M^2 = 0.75$$

$$\sigma := \sqrt{D} = 0.866$$

Відповідь: $M(X)=1{,}5$; $D(X)=\frac{3}{4}$; $\sigma(X)=\frac{\sqrt{3}}{2}$.

7.2. Показниковий закон розподілу

Означення. Неперервна випадкова величина X має *показниковий закон розподілу* з параметром λ, якщо її щільність розподілу ймовірностей має вигляд

$$f(x)=\begin{cases}0, & \text{при } x<0;\\ \lambda e^{-\lambda x}, & \text{при } x\ge 0.\end{cases} \qquad (7.2)$$

Теорема. Функція розподілу випадкової величини X з показниковим законом розподілу має вигляд

$$F(x)=\begin{cases}0, & \text{при } x<0;\\ 1-e^{-\lambda x}, & \text{при } x\ge 0;\end{cases} \qquad (7.3)$$

а її математичне сподівання і дисперсія відповідно дорівнюють

$$M(X)=\frac{1}{\lambda};\;\; D(X)=\frac{1}{\lambda^2}. \qquad (7.4)$$

Доведення.

Дійсно, при $x<0$, $F(x)=0$;

$$\text{при } x\geq 0,\; F(x)=\int_0^x \lambda e^{-\lambda t}dt=e^{-\lambda t}\Big|_0^x=1-e^{-\lambda x};$$

математичне сподівання $M(X)=\int_0^\infty x\lambda e^{-\lambda x}dx=\lim\limits_{b\to\infty}\int_0^b x\lambda e^{-\lambda x}dx=$

$$=\lim_{b\to\infty}\left(-xe^{\lambda x}\Big|_0^b+\int_0^b \lambda e^{-\lambda x}dx\right)=\frac{1}{\lambda}.$$

Пропонується перевірити самостійно, що дисперсія $D(X)=\dfrac{1}{\lambda^2}$. •

Для показникового закону розподілу:

при знаходженні щільності розподілу використовується функція **dexp(*x*,** λ**)**, функції розподілу – **pexp(*x*,** λ**)**, де ***x*** – значення випадкової величини з параметром λ.

Приклад 7.2. Неперервна випадкова величина X розподілена за показниковим законом розподілу з параметром $\lambda=3$. Записати $f(x)$ та $F(x)$. Знайти $P(1<X<1{,}5)$; $M(X)$; $D(X)$.

Розв'язання. Згідно з формулами (7.2), (7.3):

$$f(x)=\begin{cases}0, & x<0;\\ 3e^{-3x}, & x\geq 0;\end{cases}\qquad F(x)=\begin{cases}0, & x<0;\\ 1-e^{-3x}, & x\geq 0.\end{cases}$$

Ймовірність $P(1<X<1{,}5)=F(1{,}5)-F(1)=1-e^{-4{,}5}-1+e^{-3}=e^{-3}-e^{-4{,}5}=$

$$=0{,}0483-0{,}0111=0{,}0372.$$

З формул (7.4):

$$M(X)=\frac{1}{3};\; D(X)=\frac{1}{16}.$$

$$\lambda := 3$$

$$M := \int_{-\infty}^{\infty} x \cdot \mathrm{dexp}(x,\lambda)\, dx = 0.333$$

$$D := \int_{-\infty}^{\infty} x^2 \cdot \mathrm{dexp}(x,\lambda)\, dx - M^2 = 0.111$$

$$P := \mathrm{pexp}(1.5,\lambda) - \mathrm{pexp}(1,\lambda) = 0.039$$

<u>Відповідь:</u> $f(x)=\begin{cases}0, & x<0;\\ 3e^{-3x}, & x\ge 0;\end{cases}$ $\quad F(x)=\begin{cases}0, & x<0;\\ 1-e^{-3x}, & x\ge 0;\end{cases}$

$$M(X)=\frac{1}{3};\; D(X)=\frac{1}{16}.$$

7.3. Нормальний закон розподілу

Нормальний закон розподілу найбільш поширений на практиці. Головна його особливість полягає у тому, що він є граничним законом, до якого наближаються інші закони розподілу при типових умовах.

<u>*Означення.*</u> Неперервна випадкова величина X має *нормальний закон розподілу* (закон Гаусса) з параметрами a і σ, якщо її щільність розподілу ймовірностей має вигляд

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}, \tag{7.5}$$

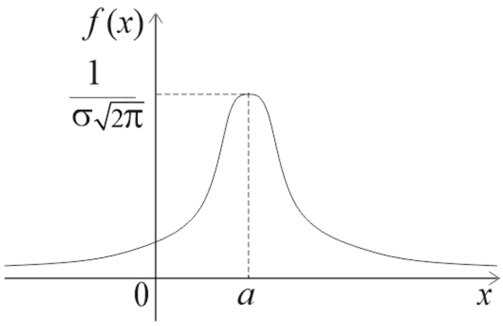
а її інтеграл

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(t-a)^2}{2\sigma^2}}dt \tag{7.6}$$

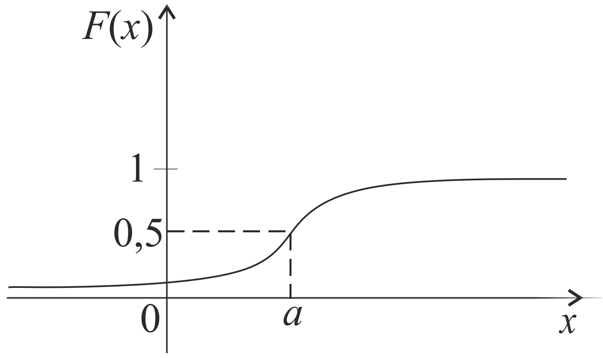
називається *нормальною функцією розподілу*.

Нормальний закон розподілу з параметрами a і σ позначається $N(a,\sigma)$.

Графіки функцій $f(x)$ і $F(x)$ для $N(a,\sigma)$. зображені на рис. 7.2 (а, б відповідно):



а)



б)

Рис. 7.2

Теорема. Якщо неперервна випадкова величина X розподілена за нормальним законом, то її математичне сподівання дорівнює параметру a, а середнє квадратичне відхилення дорівнює σ.

Доведення.

Математичне сподівання величини X: $M(X)=\int\limits_{-\infty}^{+\infty} x\cdot\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}dx.$ Нехай $t=\frac{x-a}{\sigma\sqrt{2}}$, тоді $x=a+\sigma\sqrt{2}t$ і $dx=\sigma\sqrt{2}dt.$ Отже,

$$M(X)=\int\limits_{-\infty}^{+\infty}(a+\sigma\sqrt{2}t)\cdot\frac{1}{\sigma\sqrt{2\pi}}e^{-t^2}\sigma\sqrt{2}dt=\frac{\sigma\sqrt{2}}{\sqrt{\pi}}\int\limits_{-\infty}^{+\infty}te^{-t^2}dt+\frac{a}{\sqrt{\pi}}\int\limits_{-\infty}^{+\infty}e^{-t^2}dt=$$

$$=0+\frac{a}{\sqrt{\pi}}\sqrt{\pi}=a.$$

(перший інтеграл дорівнює нулю, як інтеграл від непарної функції по симетричному відносно початку координат проміжку, а другий $\int\limits_{-\infty}^{+\infty}e^{-t^2}dt=\sqrt{\pi}$ – інтеграл Ейлера-Пуассона).

Для знаходження дисперсії потрібно зробити аналогічну заміну і застосувати метод інтегрування частинами:

$$D(X)=\int_{-\infty}^{+\infty}(x-a)^2\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}dx=\int_{-\infty}^{+\infty}\sigma^2 2t^2\frac{1}{\sigma\sqrt{2\pi}}e^{-t^2}\sigma\sqrt{2}dt=\frac{2\sigma^2}{\sqrt{\pi}}\int_{-\infty}^{+\infty}t^2e^{-t^2}dt=$$

$$=-\frac{\sigma^2}{\sqrt{\pi}}\int_{-\infty}^{+\infty}tde^{-t^2}=-\frac{\sigma^2}{\sqrt{\pi}}\left(te^{-t^2}\Big|_{-\infty}^{+\infty}-\int_{-\infty}^{+\infty}e^{-t^2}dt\right)=0+\frac{\sigma^2}{\sqrt{\pi}}\cdot\sqrt{\pi}=\sigma^2.$$

Отже, $\sigma(X)=\sqrt{D(X)}=\sigma.$ •

<u>*Зауваження.*</u> Якщо випадкова величина X розподілена за нормальним законом з параметрами $a=0,\ \sigma=1,$ тобто $N(0;1)$, то такий розподіл називається *стандартним*.

Функція щільності розподілу ймовірностей у цьому випадку буде функція Гаусса $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$, графік якої симетричний відносно осі Oy, а функція розподілу:

$$F(x)=\frac{1}{2}+\Phi(x)=\frac{1}{2}+\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{x^2}{2}}dx. \qquad (7.7)$$

Цей інтеграл не виражається через елементарні функції, тому значення функції Лапласа знаходять за допомогою таблиць (додаток 2).

Далі буде встановлено зв'язок між функцією розподілу $F(x)$ випадкової величини X, що має нормальний закон розподілу і функцією Лапласа $\Phi(x)$.

Згідно з рівністю (7.6): $F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(t-a)^2}{2\sigma^2}}dt.$

Нехай $z=\frac{t-a}{\sigma}$, тоді $t=a+z\sigma,\ t\to-\infty,\ z\to-\infty$, отже,

$$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\frac{x-a}{\sigma}}e^{-\frac{z^2}{2}}dz=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{0}e^{-\frac{z^2}{2}}dz+\int_{0}^{\frac{x-a}{\sigma}}e^{-\frac{z^2}{2}}dz.$$

Перший інтеграл знаходиться, використовуючи інтеграл Ейлера-Пуасона:

$$\int_{-\infty}^{0}e^{-\frac{t^2}{2}}dt=\frac{1}{2}\int_{-\infty}^{+\infty}e^{-\frac{t^2}{2}}dt=\frac{1}{2}\sqrt{2}\int_{-\infty}^{+\infty}e^{-\left(\frac{t}{\sqrt{2}}\right)^2}d\left(\frac{t}{\sqrt{2}}\right)=\frac{1}{2}\sqrt{2}\sqrt{\pi}=\sqrt{\frac{\pi}{2}}.$$

Другий інтеграл є $\Phi\left(\frac{x-a}{\sigma}\right)$. Отже,

$$F(x)=\frac{1}{2}+\Phi\left(\frac{x-a}{\sigma}\right). \qquad (7.8)$$

Для нормального закону розподілу:

при знаходженні щільності розподілу використовується функція **dnorm (*x*,*m*,*s*)**, функції розподілу – **pnorm(*x*,*m*,*s*)**, де ***x*** – значення випадкової величини, ***m*** – математичне сподівання, ***s*** – середнє квадратичне відхилення.

7.4. Ймовірність попадання значення нормально розподіленої випадкової величини в інтервал. «Правило трьох сигм»

Ймовірність того, що випадкова величина набуває значення з інтервалу $(\alpha;\beta)$ обчислюється за формулою

$$P(\alpha<X<\beta)=F(\beta)-F(\alpha),$$

де $F(x)$ – функція розподілу випадкової величини X.

Якщо випадкова величина розподілена за нормальним законом, то з рівності (7.4): $P(\alpha<X<\beta)=\frac{1}{2}+\Phi\left(\frac{\beta-a}{\sigma}\right)-\frac{1}{2}-\Phi\left(\frac{\alpha-a}{\sigma}\right)$, звідси

$$P(\alpha<X<\beta)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right). \qquad (7.9)$$

Геометрично функція Лапласа $\Phi(x)$ є площею фігури, що розміщена під кривою Гаусса на відрізку $[0;x]$ (рис. 7.3).

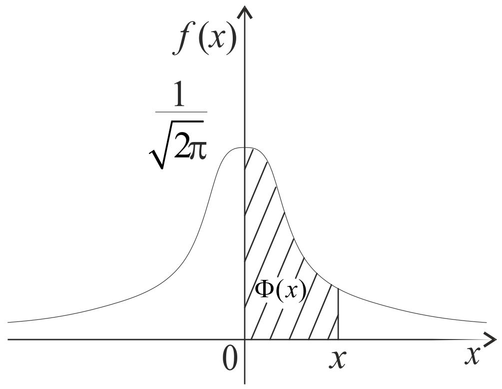


Рис. 7.3

Через функцію Лапласа можна виразити ймовірність знаходження значень випадкової величини на проміжку довжиною 2ε, симетричному щодо $M(X) = a$:

$$P(|X-a|<\varepsilon) = P(a-\varepsilon < X < a+\varepsilon) = \Phi\left(\frac{a+\varepsilon-a}{\sigma}\right) - \Phi\left(\frac{a-\varepsilon-a}{\sigma}\right) =$$

$$= \Phi\left(\frac{\varepsilon}{\sigma}\right) - \Phi\left(-\frac{\varepsilon}{\sigma}\right) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$ Отже,

$$P(|X-a|<\varepsilon) = 2\Phi\left(\frac{\varepsilon}{\sigma}\right).$$

Якщо $\varepsilon = 3\sigma,$ то

$$P(|X-a|<3\sigma) = 2\Phi(3) \approx 0{,}9973.$$

Тоді

$$P(|X-a|\geq 3\sigma) = 1 - P(|X-a|<3\sigma) = 1 - 2\Phi(3) \approx 1 - 0{,}9973 = 0{,}0027.$$

Події з ймовірністю появи 0,9973 вважаються практично вірогідними. Звідси отримується «правило трьох сигм».

Якщо випадкова подія X має нормальний закон розподілу з параметрами a і σ тобто $N(a,\sigma),$ то практично вірогідно, що всі її значення зосереджені у проміжку

$$a - 3\sigma < X < a + 3\sigma.$$

<u>Приклад 7.3.</u> Вважається, що зріст чоловіків певної вікової групи є нормально розподілена випадкова величина X з пераметрами $a = 173$, $\sigma^2 = 36$. Знайти частку костюмів 4-го росту (176 – 182) та 3-го росту (170 – 176), які потрібно передбачити у загальному обсязі виробництва.

<u>Розв'язання.</u> Частка костюмів 4-го росту у загальному обсязі виробництва визначається за формулою (7.9) як ймовірність

$$P(176 \le X \le 182) = \Phi\left(\frac{182-173}{6}\right) - \Phi\left(\frac{176-173}{6}\right) = \Phi(1{,}5) - \Phi(0{,}5) =$$

$$= 0{,}4332 - 0{,}1915 = 0{,}2417\,.$$

Оскільки $a = M(X) = 173$ є середина інтервалу $(170;176)$, то нерівності $170 \le X \le 182$ рівносильна нерівність $|X - 173| < 3$.

Отже, $P\left(|X - 173| < 3\right) = 2\Phi\left(\dfrac{3}{6}\right) = 2\Phi(0{,}5) = 0{,}3831$.

```
a := 173   σ := 6
P1 := pnorm(182,a,σ) − pnorm(176,a,σ) = 0.242
P2 := pnorm(176,a,σ) − pnorm(170,a,σ) = 0.383
```

<u>Відповідь:</u> частка костюмів, які потрібно передбачити у загальному обсязі виробництва, 4-го росту (176 – 182) складає близько 24 % та 3-го росту (170 – 176) – 38 %.

7.5. Закони розподілів χ^2, Стьюдента та Фішера-Снедекора

Далі буде розглянуто деякі основні закони розподілу випадкових величин, які є функціями від нормально розподілених випадкових величин. Ці закони в подальшому будуть складати необхідний апарат для побудови статистичних критеріїв і оцінок, що застосовуються у математичній статистиці.

χ^2 *- розподіл.*

<u>*Означення.*</u> *Розподіл* χ^2 *(хі-квадрат) з* k *ступенями свободи* називається розподіл суми квадратів k незалежних випадкових величин, що мають стандартний нормальний розподіл, тобто

$$\chi^2 = \sum_{i=1}^{k} z_i^2,$$

де z_i $(i = \overline{1,k})$ має розподіл $N(0;1)$.

Щільність розподілу ймовірностей χ^2 розподілу виражається через γ (гама) функцію Ейлера (її вигляд наводитись не буде). На рис. 7.4 зображено графіки щільностей розподілу для деяких ступенів свободи k.

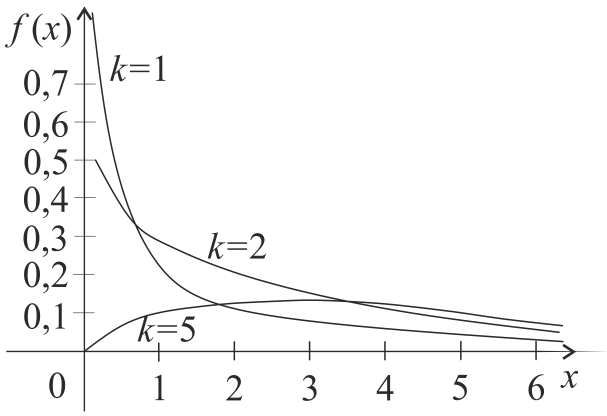


Рис. 7.4

<u>*Зауваження.*</u> При $k > 30$ розподіл випадкової величини $Z = 2\chi^2 - \sqrt{2k}$ близький до стандартного закону $N(0;1)$.

Розподіл Стьюдента.

<u>*Означення.*</u> *Розподілом Стьюдента* (або t – *розподілом*) називається розподіл випадкової величини

$$t = \frac{Z}{\sqrt{\frac{1}{k}\chi^2}},$$

де Z – випадкова величина зі стандартним законом розподілу, тобто $N(0;1)$,

χ^2– незалежна від Z випадкова величина, що має χ^2 розподіл з k ступенями свободи.

На рис. 7.5 наведена крива розподілу Стьюдента.

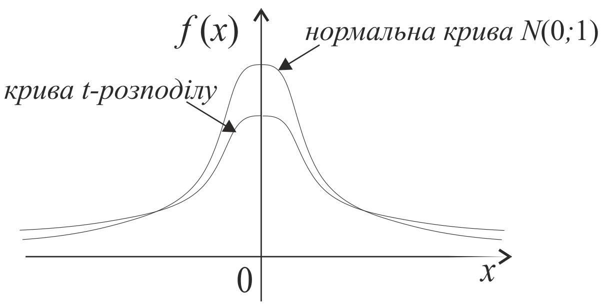


Рис. 7.5

Крива t – розподілу симетрична відносно осі ординат і при $k \to +\infty$ наближається до нормального розподілу. На практиці вже при $k > 30$ t – розподіл замінюють нормальним розподілом.

Розподіл Фішера-Снедекора.

Означення. Розподілом Фішера-Снедекора (або F – розподілом) називається розподіл випадкової величини

$$F = \frac{\frac{1}{k_1}\chi^2(k_1)}{\frac{1}{k_2}\chi^2(k_2)},$$

де $\chi^2(k_1)$ і $\chi^2(k_2)$ – випадкові величини, що мають χ^2 розподіл з k_1 і k_2 ступенями свободи.

При $k_1 \to +\infty$ і $k_2 \to +\infty$ F – розподіл наближається до нормального.

Для законів розподілу χ^2, Стьюдента та Фішера-Снедекора:

при знаходженні щільності розподілу використовуються функції:

dchisq (*x*,*d*) – для розподілу χ^2;

dt (*x*,*d*) – для розподілу Стьюдента;

dF (*x*,*d*) – для розподілу Фішера-Снедекора;

функції розподілу:

pchisq (*x*,*d*) – для розподілу χ^2;

pt (*x*,*d*) – для розподілу Стьюдента;

pF (*x*,*d₁*, *d₂*) – для розподілу Фішера-Снедекора;

де ***x*** – значення випадкової величини, $d > 0$, $d_1 > 0$, $d_2 > 0$ – ступені свободи.

У лекціях 4 – 7 було розглянуто основні закони розподілу дискретних і неперервних випадкових величин. У таблиці 7.1 подано відомості про ці закони та їхні числові характеристики.

Таблиця 7.1

Основні закони розподілу дискретної випадкової величини			
Закон розподілу X	$M(X)$	$D(X)$	$\sigma(X)$
1. Біноміальний $P(X = m) = C_n^m p^m q^{n-m}$, $q = 1 - p$, $m = \overline{0, n}$.	np	npq	$\sqrt{npq}$
2. Пуассона $P(X = m) = \dfrac{\lambda^m}{m!} e^{-\lambda}$, $m = 0, 1, 2 \ldots$, $\lambda > 0$.	λ	λ	$\sqrt{\lambda}$
3. Геометричний $P(X = m) = pq^{m-1}$, $m = 0, 1, 2 \ldots$, $q = 1 - p$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$	$\dfrac{\sqrt{q}}{p}$

Основні закони розподілу неперервної випадкової величини			
Закон розподілу X	$M(X)$	$D(X)$	$\sigma(X)$
1. Рівномірний $F(x)=\begin{cases}0, & \text{при } x\le a;\\ \dfrac{1}{b-a}, & \text{при } a<x\le b;\\ 1, & \text{при } x>b.\end{cases}$	$\dfrac{a+b}{2}$	$\dfrac{(a-b)^2}{12}$	$\dfrac{b-a}{2\sqrt{3}}$
2. Показниковий $F(x)=\begin{cases}0, & \text{при } x<0;\\ \lambda e^{-\lambda x}, & \text{при } x\ge 0.\end{cases}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	$\dfrac{1}{\lambda}$
3. Нормальний $F(x)=\dfrac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}.$	a	σ^2	σ

Практичне заняття 7

7.1. Випадкова величина X розподілена рівномірно в інтервалі $(1;7)$. Знайти: 1) функцію розподілу $F(x)$ та щільність розподілу $f(x)$ випадкової величини X, побудувати їх графіки; 2) математичне сподівання і дисперсію випадкової величини X; 3) ймовірність попадання випадкової величини X в інтервал $(1;9)$.

7.2. Випадкова величина X розподілена рівномірно в інтервалі $(3;9)$. Знайти: 1) функцію розподілу $F(x)$ та щільність розподілу $f(x)$ випадкової величини X, побудувати їх графіки; 2) математичне сподівання і дисперсію випадкової величини X; 4) ймовірність попадання випадкової величини X в інтервал $(0;4)$.

7.3. Випускник університету влаштовується на роботу. Його сповістили, що для співбесіди з ним зв'яжуться між 10 і 12 годинами дня. Яка ймовірність того, що початок співбесіди йому прийдеться чекати не більше 15 хвилин?

7.4. Марина і Юлія придбали пакет соку, який має об'єм 2 л. Марина випила частину соку. Яка ймовірність, що Юлії залишилось не менше 1 л соку?

7.5. Поїзди метро їздять згідно розкладу, з інтервалом 5 хв. Знайти: 1) функцію розподілу $F(x)$ та щільність розподілу $f(x)$ випадкової величини X – часу очікування чергового поїзда, побудувати їх графіки; 2) математичне сподівання і дисперсію випадкової величини X.

7.6. Автобуси маршруту Київ-Житомир їздять строго за розкладом, з інтервалом у 30 хв. Знайти ймовірність того, що пасажир, який підійшов до зупинки, буде чекати черговий автобус не більше 10 хв.?

7.7. Стрілка кварцового годинника переміщується стрибком у кінці кожної хвилини. Знайти ймовірність того, що у дану мить годинник покаже час, який відрізнятиметься від реального не більше, ніж на 30 с.

7.8. Відомо, що випадкова величина X розподілена за показниковим законом розподілу із параметром $\lambda = 2$. Знайти: 1) щільність розподілу та функцію розподілу випадкової величини X; 2) математичне сподівання і дисперсію випадкової величини X.

7.9. Неперервна випадкова величина X розподілена за показниковим законом розподілу із параметром $\lambda = 5$. Знайти: 1) щільність розподілу та функцію розподілу випадкової величини X; 2) математичне сподівання і дисперсію випадкової величини X; 3) ймовірність попадання випадкової величини X в інтервал $(0,5;1)$.

7.10. Знайти щільність розподілу та функцію розподілу випадкової величини X, якщо відомо, що вона має показниковий розподіл, її математичне сподівання дорівнює 0,4.

7.11. Тривалість безвідмовної роботи елемента має показниковий розподіл з параметром $\lambda = 0,01$. Знайти ймовірність того, що за час $t = 60$ год: 1) елемент відмовить; 2) елемент не відмовить.

7.12. Відомо, що 99 % моторів, якими оснащене фасувальне обладнання підприємства по виробництву соняшникової олії та майонезів, виходить із ладу після 5000 мотогодин. Яка ймовірність того, що мотор вийде з ладу в інтервалі часу від 3000 до 4000 мотогодин. Оцінити середній час безвідмовної роботи моторів і дисперсію. *Вказівка.* Припустити, що неперервна випадкова величина «час безвідмовної роботи мотора» має показниковий розподіл.

7.13. У деякому казковому королівстві Нухтія було проведене дослідження, яке показало, що ймовірність його мешканцю дожити до 100 років становить 0,35. Яка ймовірність того, що новонароджений син короля доживе до 21 року?

7.14. Час очікування у черзі до банкомату у період отримання стипендії має показниковий закон розподілу із середнім часом очікування 15 хв. Яка ймовірність того, що студент витратить на очікування не менше 5 і не більше 10 хв.?

7.15. Час виконання замовлення на випічку весільного короваю в середньому становить 2 доби. Яка ймовірність того, що замовлений молодятами коровай буде випечений не раніше, ніж через 1 добу?

7.16. На олійно-жировому комбінаті випробовуються дві сушарки для сушки насіння соняшника. Продовжуваність безвідмовної роботи першої має показниковий розподіл $F_1(t)=1-e^{-0,03t}$, другої – $F_2(t)=1-e^{-0,05t}$. Знайти ймовірність того, що за добу: 1) відмовить хоча б одна сушарка; 2) відмовить тільки одна сушарка; 3) відмовлять обидві сушарки; 4) обидві сушарки будуть працювати безвідмовно.

7.17. Працюючи без технічної перерви оператор банку протягом години в середньому робить 3 помилки. Оцінити ймовірність того, що за три години роботи без технічної перерви оператор зробить: 1) хоча б одну помилку; 2) одну помилку; 3) 5 помилок; 4) не більше 4 помилок.

7.18. Нормально розподілена випадкова величина має параметри $a=10$, $\sigma=12$. Потрібно: 1) скласти функцію щільності розподілу; 2) знайти ймовірність того, що випадкова величина набуде значень з інтервалу $(10;17)$.

7.19. Випадкова величина розподілена нормально з параметрами $a = 50$, $\sigma = 5$. Потрібно: 1) скласти функції розподілу та щільності розподілу; 2) знайти ймовірність того, що випадкова величина набуде значення більшого за $\alpha = 39$ та меншого за $\beta = 55$; 3) знайти ймовірність того, що значення випадкової величини відхилиться від середнього значення не більше, ніж на $\varepsilon = 3$.

7.20. Нехай кількість реалізованих за тиждень тортів «Ромашка» – X, нормально розподілена випадкова величина, її математичне сподівання $M(X) = 350$ шт. Середнє квадратичне відхилення цієї випадкової величини $\sigma = 10$ шт. Оцінити ймовірність того, що за тиждень буде продано від 345 до 380 шт. тортів.

7.21. Дослідженням встановлено, що кількість клієнтів, які звертаються у страхову компанію «Скай» за квартал – X, нормально розподілена випадкова величина з математичним сподіванням $M(X) = 1{,}5$ тис. чоловік та середнім квадратичним відхиленням $\sigma = 0{,}05$ тис. чоловік. Оцінити ймовірність того, що за квартал у компанію «Скай» звернеться не менше 1 і не більше 1,2 тис. чоловік.

7.22. Відомо, що на молокозаводі добове споживання електроенергії є нормально розподіленою випадковою величиною, математичне сподівання якої 5000 кВт/год, дисперсія 300 кВт/год. Оцінити ймовірність того, що за найближчу добу споживання електроенергії на підприємстві становитиме від 3000 до 4500 кВт/год.

7.23. Підприємство «Веселка» щомісячно виробляє та реалізує продукцію дитячого харчування. Кількість замовлень на сирову масу «Сонечко» є нормально розподіленою випадковою величиною з середнім квадратичним відхиленням 380. Відомо, що у 95 % випадків кількість щомісячних замовлень перевищує 9827. Знайти середню кількість замовлень, які отримує підприємство за місяць?

7.24. На кондитерській фабриці встановлено новий автомат, який нарізає східні солодощі довжиною 5 см. Відхилення від заданої довжини вважається стандартним, якщо воно не перевищує 0,5 см. Випадкові відхилення розподілені нормально із математичним сподіванням 4,9 см, та середнім квадратичним відхиленням 0,45 см. Скільки відсотків східних солодощів із стандартною довжиною нарізає автомат?

7.25. Ціна акції підприємства «Молочні ріки» розподілена за нормальним законом розподілу з математичним сподіванням 100 ум. од. та середнім квадратичним відхиленням 5 ум. од. Знайти межі, в яких буде знаходитись ціна акцій.

7.26. За даними бухгалтерії заробітна плата по підприємству має нормальний розподіл із середнім значенням 4500 грн. та середнім квадратичним відхиленням 1500 грн. Знайти ймовірність того, що випадково вибраний працівник підприємства отримує заробітну плату: 1) меншу 9000 грн.; 2) від 2800 до 3500 грн.

7.27. Ціна акції пивзаводу «Козачок» розподілена за нормальним законом розподілу. Протягом останнього року п'яту його частину ціна була менша за 15 ум. од., а 65 % днів вона була більша за 20 ум. од. Знайти математичне сподівання та середнє квадратичне відхилення ціни акції.

7.28. Вага тістечок «Колобок», які фасуються у картонні коробки, нормально розподілена випадкова величина. Відомо, що 75 % коробок мають вагу більшу, ніж 50 г, 15 % – меншу, ніж 45 г. Знайти середню вагу та середнє квадратичне відхилення ваги коробки.

7.29. На телефонний номер служби доставки ресторану «Первачок» від замовників поступає 40 дзвінків на добу. Кожний замовник витрачає у середньому 5 хв. на замовлення. Скільки каналів повинен мати багатоканальний зв'язок, щоб із практично вірогідною ймовірністю, можна було стверджувати, що ресторан не втратить жодного замовника?

7.30. На підприємстві «Водограй» зупинилась автоматизована лінія по розливу соків. Інженер встановив, що розміри заміненої нещодавно деталі не відповідають допустимим, тобто виходять за межі допустимого інтервалу (29 мм; 32 мм). Відомо, що розмір цих деталей нормально розподілена випадкова величина, яка має математичне сподівання 30 мм та середнє квадратичне відхилення 2 мм. Оцінити ймовірність того, що причиною зупинки автоматизованої лінії стали нестандартні розміри деталі.

Лекція 8. Закони великих чисел

8.1. Нерівності Чебишова

8.2. Теореми Чебишова

8.3. Теорема Бернуллі. Центральна гранична теорема

Як відомо, неможливо наперед впевнено передбачити, яке з можливих значень набуде випадкова величина в результаті експерименту. Але, коли проводиться велика кількість випробувань, то при певних умовах, характеристики випадкових величин стають майже невипадковими. Наприклад, частота події при великій кількості випробувань стає стійкою, те ж саме стосується і середнього значення випадкової величини. Ці факти дозволяють використати результати спостережень над випадковими явищами для передбачення результатів майбутніх випробувань.

Теореми, в яких за певних умов встановлюється факт наближення середніх характеристик випадкових величин під час проведення великої кількості експериментів до певних сталих величин, об'єднують під спільною назвою «закони великих чисел».

Перед тим, як перейти до цих теорем, потрібно розглянути нерівності Маркова і Чебишова, які потрібні для їх доведення.

8.1. Нерівність Чебишова

Лема Чебишова (нерівність Маркова). Якщо випадкова величина X набуває тільки невід'ємних значень і має скінченне математичне сподівання, тоді для довільного додатного числа ε

$$P(X > \varepsilon) \le \frac{M(X)}{\varepsilon}. \qquad (8.1)$$

Доведення. Доведення приведене для дискретної випадкової величини X. Нехай значення випадкової величини розташовані у порядку зростання так, що значення $x_1, x_2, ..., x_k$ не більші за число ε, а інші значення $x_{k+1}, ..., x_n$ будуть більші за ε. Тоді математичне сподівання $M(X)$:

$$M(X) = x_1 p_1 + x_2 p_2 + ... + x_k p_k + x_{k+1} p_{k+1} + ... + x_n p_n.$$

Відкинувши перші k невід'ємних доданків, можна отримати нерівність:

$$x_k p_k + x_{k+1} p_{k+1} + ... + x_n p_n \le M(X). \qquad (8.2)$$

В нерівності (8.2) всі значення $x_i,\ i = \overline{k+1, n}$, більші за ε, отже,

$$\varepsilon(p_k + p_{k+1} + ... + p_n) \le M(X)$$

або

$$p_k + p_{k+1} + ... + p_n \le \frac{M(X)}{\varepsilon}.$$

Сума ймовірностей у лівій частині останньої нерівності є ймовірністю того, що випадкова величина X набуде значення більшого ε: $P(X > \varepsilon)$.

Тому $P(X > \varepsilon) \le \dfrac{M(X)}{\varepsilon}$. •

Користуючись тим, що події $X > \varepsilon$ та $X \le \varepsilon$ протилежні, можна отримати:

$$P(X \le \varepsilon) \ge 1 - \frac{M(X)}{\varepsilon}. \qquad (8.3)$$

<u>Приклад 8.1.</u> Середня щільність телефонних викликів на комутатор швидкої допомоги протягом години становить 90. Оцінити ймовірність того, що протягом наступної години число викликів на комутатор: а) буде більше 200; б) буде не більше 300.

<u>Розв'язання.</u> а) за умовою $M(X) = 90$. З формули (8.1):

$$P(X > 200) \le \frac{90}{200} = 0,45,$$

тобто ймовірність того, що число викликів перевищить 200, буде не більша 0,45.

б) З формули (8.3):

$$P(X \le 300) \ge 1 - \frac{90}{300} = 0,7.$$

Отже, ймовірність того, що число викликів не більше 300, буде не менша 0,7.

Відповідь: ймовірність буде а) не більша 0,45; б) не менша 0,7.

Теорема 8.1. (Нерівність Чебишова). Якщо випадкова величина X має скінчену дисперсію $D(X)$, то для довільного $\varepsilon > 0$ виконується нерівність:

$$P(|X - M(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}. \qquad (8.4)$$

Доведення. Застосуємо нерівність (8.1) у випадку, коли $X_1 = (X - M(X))^2$ і $\varepsilon_1 = \varepsilon^2$. Тоді

$$P\left((X - M(X))^2 \geq \varepsilon_1\right) \leq \frac{M(X - M(X))^2}{\varepsilon_1}.$$

Нерівність $(X - M(X))^2 \geq \varepsilon_1$ рівносильна нерівності $|X - M(X)| \geq \varepsilon,$ а $M(X - M(X))^2$ є дисперсією випадкової величини X, отже,

$$P(|X - M(X)| \geq \varepsilon) \leq \frac{D(X)}{\varepsilon^2}. \bullet$$

Враховуючи, що події $|X - M(X)| \geq \varepsilon$ та $|X - M(X)| < \varepsilon$ – протилежні, нерівність Чебишова (8.4) можна переписати у вигляді:

$$P(|X - M(X)| < \varepsilon) \geq 1 - \frac{D(X)}{\varepsilon^2}. \qquad (8.5)$$

Приклад 8.2. Ймовірність того, що з конвеєра виходить стандартна деталь 0,95. Оцінити ймовірність того, що число бракованих деталей серед 1000 буде не більше 60 і не менше 40.

Розв'язання. Нехай $X = m$ – випадкова величина числа бракованих деталей. Ця величина має біноміальний розподіл, тому $M(X) = np$ і $D(X) = npq$. Із умови задачі $n = 1000$, $q = 0,95$ і $p = 1 - 0,95 = 0,05$. Отже, $M(X) = 50$, $D(X) = 47,5$.

Ймовірність $P(40 < X < 60)$ можна записати, як $P(-10 < X - 50 < 10)$ або $P(|X - 50| < 10)$. Підставивши у формулу (8.5) $\varepsilon=10$, $M(X) = 50$, $D(X) = 47{,}5$, можна отримати:

$$P(|M(X) - 50| < 10) \geq 1 - \frac{47{,}5}{100},$$

$$P(|M(X) - 50| < 10) \geq 0{,}525.$$

<u>Відповідь:</u> ймовірність буде не менша 0,525.

8.2. Теореми Чебишова

<u>Теорема 8.2.</u> Якщо випадкові величини $X_1, X_2, ..., X_n$ попарно незалежні, мають дисперсії, обмежені в сукупності однією і тією ж сталою, тобто $D(X_i) \leq C \quad (i = \overline{1,n})$. Тоді для довільного $\varepsilon > 0$

$$\lim_{n\to\infty} P\left(\left| \frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n} M(X_k) \right| < \varepsilon \right) = 1. \tag{8.6}$$

<u>Доведення.</u> Нехай $\bar{X}_n$ середнє значення випадкової величини $X_1, X_2, ..., X_n$: $\bar{X}_n = \frac{1}{n}\sum_{k=1}^{n} X_k$. Тоді $M\left(\bar{X}_n\right) = \frac{1}{n}\sum_{k=1}^{n} M(X_k)$, $D\left(\bar{X}_n\right) = \frac{1}{n^2}\sum_{k=1}^{n} D(X_k)$.

Потрібно розглянути $P(|\bar{X}_n - M(\bar{X}_n)| < \varepsilon) = 1 - P(|\bar{X}_n - M(\bar{X}_n)| \geq \varepsilon)$. З нерівності Чебишова і властивості дисперсій можна отримати наступну оцінку:

$$P(|\bar{X}_n - M(\bar{X}_n)| < \varepsilon) = 1 - P(|\bar{X}_n - M(\bar{X}_n)| \geq \varepsilon) \geq 1 - \frac{1}{\varepsilon^2}\frac{1}{n^2}\sum_{k=1}^{n} D(\bar{X}_k) \leq 1 - \frac{C}{\varepsilon^2 n}. \tag{8.7}$$

Перейшовши в останній нерівності до границі, можна отримати, що

$$\lim_{n\to\infty} P\left(\left|\bar{X}_n - M\left(\bar{X}_n\right)\right| < \varepsilon\right) \geq 1$$

або

$$\lim_{n\to\infty} P\left(\left| \frac{1}{n}\sum_{k=1}^{n} X_k - \frac{1}{n} M(X_k) \right| < \varepsilon \right) = 1.$$

Звідси і з того, що ймовірність не може бути більшою за 1 встановлюється справедливість твердження теореми Чебишова.•

Наслідок. Якщо випадкові величини $X_1, X_2, ..., X_n$ попарно незалежні, мають однакові математичні сподівання $M(X_i) = a,\ \ i = \overline{1,n}$ і їх дисперсії обмежені однією і тією ж сталою, тоді $\lim\limits_{n\to\infty} P\left(\left|\frac{1}{n}\sum\limits_{k=1}^{n} X_k - a\right| < \varepsilon\right) = 1.$

З теореми Чебишова випливає, що середнє арифметичне випадкових величин $X_1, X_2, ..., X_n$ для досить великих n можна замінити на середнє арифметичне їх математичних сподівань, тобто для досить великих n з будь-якою точністю має місце наближена рівність

$$\frac{1}{n}\sum_{k=1}^{n} X_k \approx \frac{1}{n}\sum_{k=1}^{n} M(X_k).$$

Приклад 8.3. Скільки потрібно провести вимірювань певної величини, щоб із ймовірністю не менше 0,96 гарантувати відхилення цих вимірювань від істинного значення величини не більше, ніж на 1 (за абсолютною величиною), якщо середнє квадратичне відхилення кожного з них не перевищує 4?

Розв'язання. Нехай $X_i,\ \ i = \overline{1,n}$ – результат $i-$го вимірювання, дійсне значення величини $M(X_i) = a$ при будь-якому i, $i = \overline{1,n}$. Необхідно знайти n, для якого

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{k=1}^{n} X_k - a\right| \le 1\right) \ge 0{,}96.$$

З формули (8.7):

$$1 - \frac{C}{\varepsilon^2 n} = 1 - \frac{4^2}{1^2 n} > 0{,}96 \Rightarrow \frac{4^2}{n} \le 0{,}04 \Rightarrow n \ge \frac{16}{0{,}04} = 400.$$

Відповідь: потрібно не менше 400 вимірювань.

8.3. Теорема Бернуллі. Центральна гранична теорема

Теорема 8.3. (Бернуллі). Нехай m – число появ події A в n незалежних випробуваннях. Тоді для довільного $\varepsilon > 0$ виконується рівність

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) = 1. \tag{8.8}$$

Ця теорема є наслідком теореми Чебишова для $\bar{X} = \frac{m}{n}$, $M(\bar{X}) = \frac{1}{n}M(m) = \frac{1}{n}np = p,\ D(\bar{X}) = \frac{1}{n^2}D(m) = \frac{pq}{n}.$

Рівність (8.8) показує, що відносна частота $\frac{m}{n}$ події A в n незалежних випробуваннях наближається до її ймовірності появи в кожному випробуванні p.

Під назвою «центральна гранична теорема» об'єднується група теорем, які стосуються граничних законів розподілу. З них випливає важливий факт, який є дуже корисним для практичного застосування: якщо $X_1, X_2, ..., X_n$ – незалежні випадкові величини, внесок кожної з яких в їхню загальну суму незначний, та, які мають той самий закон розподілу зі скінченою дисперсією, то за необмеженого зростання n закон розподілу суми $Y = \sum_{i=1}^{n} X_i$ необмежено наближається до нормального.

Серед «центральних граничних теорем» найважливішою є наступна теорема.

Теорема Ляпунова. Якщо $X_1, X_2, ..., X_n$ – незалежні випадкові величини, кожна з яких має математичне сподівання $M(X_i) = a_i$, дисперсію $D(X_i) = \sigma_i^2$ та $M\left(\left|X_i - a_i\right|^3\right) = m_i$, а також $\lim\limits_{n\to\infty} \frac{\sum_{i=1}^{n} m_i}{\left(\sum_{i=1}^{n}\sigma_i^2\right)^{3/2}} = 0,\ \lim\limits_{x\to\infty} \frac{\sum_{i=1}^{n} m_i}{\sum_{i=1}^{n}\sigma_i^2} = 0,$ тоді закон розподілу суми $Y_n = X_1 + X_2 + ... + X_n$ при $n \to +\infty$ необмежено наближується до нормального з математичним сподіванням $\sum_{i=1}^{n} a_i$ і дисперсією $\sum_{i=1}^{n}\sigma_i^2$.

Практичне заняття 8

8.1. Студенти вирішили для новорічного свята зробити гірлянду із 300 лампочок. Ймовірність того, що лампочка горітиме становить 0,7. Оцінити ймовірність того, що горітиме не менше 250 і не більше 280 лампочок.

8.2. Місячні витрати води на хлібокомбінаті «Батон» складають 300 тис. $м^3$. Оцінити ймовірність того, що найближчої місяця витрати води не перевищать 500 тис. $м^3$.

8.3. Фермер розрахував, що середня вага капустини у цьогорічному його врожаї становить 4 кг. Оцінити ймовірність того, що навмання придбана у цього фермера капустина важитиме не більше 7 кг.

8.4. Середнє споживання електроенергії на підприємстві за місяць становить 450 тис. кВт/год. Оцінити ймовірність того, що наступного місяця споживання електроенергії не перевищить 600 тис. кВт/год.

8.5. Банк «Кряк» містить вклади сумою 5 млн. ум. од. Ймовірність того, що випадковим чином узятий вклад буде не більше 7 тис. ум. од., дорівнює 0,75. Оцінити кількість вкладників банку.

8.6. Середні добові витрати підприємство «Кісточка» на закупку яловичини та свинини складають 10 тис. ум. од. з середнім квадратичним відхиленням 500 ум. од. Оцінити ймовірність того, що витрати на закупку м’ясної сировини наступного дня не перевищать 12 тис. ум. од.

8.7. Місткість пакету соку є величина випадкова, середнє значення якої становить 1 л. Середнє квадратичне відхилення цієї величини становить 0,02 л. Оцінити ймовірність того, що відхилення об’єму пакету соку від його середнього значення не більше 0,04 л (за абсолютною величиною).

8.8. Середньодобове споживання електроенергії на підприємстві становлять 15000 кВт/год з середнім квадратичним відхиленням 500 кВт/год. Оцінити споживання електроенергії на найближчу добу з ймовірністю не менше, ніж 0,97.

8.9. Дослідження показують, що в середньому 70 % клієнтів страхової компанії «Добрострах» оминають страховий випадок і не потребують виплат страхової суми. Оцінити ймовірність того, що із 500 клієнтів цієї страхової компанії частка тих, хто не потребуватиме виплат буде відрізнятись від ймовірності не більше, ніж на 0,05 (за абсолютною величиною).

8.10. Відділом контролю підприємства встановлено, що у середньому 10 тістечок «Сніжинка» із 100 у тій, чи іншій мірі не відповідають стандарту. Оцінити ймовірність того, що частка нестандартних тістечок серед 300 виготовлених на замовлення відрізнятиметься від математичного сподівання цієї частки не більше, ніж на 0,1 (за абсолютною величиною).

8.11. Середній час запізнення студента на лекцію складає 2 хв. Оцінити ймовірність того, що студент запізниться не менше, ніж на 5 хв.

8.12. Університетська їдальня замовила на свято 100 коробок з заварними тістечками (по 6 шт. у коробці). Середня кількість тістечок у кожній коробці, які прим’ялись при транспортуванні, становить 2 шт., середнє квадратичне відхилення – 1 шт. Знайти межі, в яких з ймовірністю не менше 0,95 буде міститись загальне число тістечок, які прим’ялись при транспортуванні.

8.13. Першокурсники вирішили визначити середній зріст хлопців їхнього курсу. Скільки потрібно випадковим чином відібрати хлопців, щоб із ймовірністю більшою за 0,98, можна було стверджувати, що середній зріст у відібраної групи студентів буде відрізнятись від середнього зросту усіх першокурсників за абсолютною величиною не більше, ніж на 1 см. Відомо, що середнє квадратичне відхилення зросту для кожного студента із відібраної групи не перевищує 5 см.

8.14. На хлібокомбінаті відділ контролю перевіряють партію випечених батонів. З ймовірністю 0,02 батон має брак, пов'язаний із неякісними інгредієнтами, які входять до складу тіста, і незалежно від цього, з ймовірністю 0,03 – брак, пов'язаний із процесом випікання. У яких межах майже напевно буде знаходитись кількість бракованих батонів, якщо перевіряється партія із 900 батонів? *Вказівка*. За ймовірність події, яка майже напевно відбудеться, прийняти 0,99.

8.15. Банк «Курінь» вирішив опитати своїх клієнтів на рахунок задоволення їх роботи із банком. Дослідженням встановлено, що з ймовірністю 0,03 клієнтів не влаштовує компетентність співробітників банку, і незалежно від цього, з ймовірністю 0,01 – відсоток, який пропонує банк по депозитному рахунку. Випадковим чином було відібрано 1000 клієнтів. У яких межах майже напевно буде знаходитись кількість невдоволених клієнтів банку? *Вказівка*. За ймовірність події, яка майже напевно відбудеться, прийняти 0,99.

8.16. Страхова компанія нараховує 4000 клієнтів. Ймовірність того, що із них 2000 звернеться із страховим випадком становить 0,3; 1500 – 0,4 та 1000 – 0,7. Знайти межі, у яких повинна міститись частота появи страхового випадку, якщо це необхідно гарантувати із ймовірністю 0,95.

Лекція 9. Системи випадкових величин

9.1. Поняття системи випадкових величин та закони їх розподілу

9.2. Функція розподілу системи двох випадкових величин

9.3. Щільність розподілу ймовірностей двох випадкових величин

9.4. Умовний закон розподілу

9.5. Незалежні випадкові величини

9.6. Числові характеристики двовимірної випадкової величини

9.1. Поняття системи випадкових величин та закони їх розподілу

На практиці результати випробування часто характеризуються не однією випадковою величиною, а двома і більше випадковими величинами, які утворюють комплекс або систему випадкових величин $X_1, X_2, ..., X_n$. Такі системи, в залежності від кількості складових компонентів, називаються двовимірними, трьохвимірними, …, $n-$ вимірними випадковими величинами.

Наприклад:

1. Успішність випускника вищого навчального закладу характеризується системою n випадкових величин: $X_1, X_2, ..., X_n$ – балами з різних дисциплін.

2. Погода в деякій місцевості у визначений момент доби характеризується системою випадкових величин: X_1 – температурою, X_2 – вологістю, X_3 – тиском, X_4 – швидкістю вітру.

Властивості системи багатьох випадкових величин визначаються не тільки властивостями окремих випадкових величин, що входять у цю систему, але і взаємними зв'язками між випадковими величинами.

Для спрощення подання матеріалу, можна зупинитись на вивченні двовимірної системи випадкових величин (X,Y).

Зауваження. Двовимірну випадкову величину можна розглядати як випадкову точку з випадковими координатами (X,Y) на площині, або як випадковий вектор з координатами (X,Y).

<u>Означення.</u> Законом розподілу двох випадкових величин (X,Y) називається співвідношення між двома можливими значеннями (X,Y) та їх ймовірностями.

<u>Означення.</u> Система випадкових величин називається *дискретною*, якщо її компоненти є дискретними випадковими величинами, і *неперервною*, якщо її компоненти є неперервними випадковими величинами.

Закон розподілу двовимірної випадкової величини можна задати у вигляді таблиці:

$X \backslash Y$	y_1	y_2	$\dots$	y_n
x_1	p_{11}	p_{12}	$\dots$	p_{1n}
x_2	p_{21}	p_{22}	$\dots$	p_{2n}
$\dots$	$\dots$	$\dots$	$\dots$	$\dots$
x_m	p_{m1}	p_{m2}	$\dots$	p_{mn}

де p_{ij} – ймовірність сумісної появи подій $(X = x_i)$ та $(Y = y_j)$ і $\sum_{i=1}^{m}\sum_{j=1}^{n} p_{ij} = 1.$

Остання рівність випливає з того, що всі події в таблиці несумісні і утворюють повну групу подій. Знаючи закон розподілу двовимірної дискретної випадкової величини, можна знайти закон розподілу кожної із її складових X і Y. Дійсно, наприклад, події $(X = x_1, Y = y_1)$, $(X = x_1, Y = y_2)$, $\dots$, $(X = x_1, Y = y_n)$ несумісні, тому ймовірність того, що X набуде значення x_1, за теоремою додавання дорівнює:

$$P(x_1) = p_{11} + p_{12} + \dots + p_{1n}.$$

Отже, ймовірність того, що $X = x_i$ визначається, як сума ймовірностей $i-$го рядка таблиці. Аналогічно, додаванням ймовірностей $j-$го стовпчика, визначається ймовірність $P(Y = y_j)$.

<u>Приклад 9.1.</u> Знайти закони розподілу складових двовимірної випадкової величини, яка задана законом розподілу

$X \backslash Y$	y_1	y_2	y_3
x_1	0,15	0,25	0,05
x_2	0,1	0,25	0,2

<u>Розв'язання.</u>

$P(X = x_1) = 0{,}15 + 0{,}25 + 0{,}05 = 0{,}45$,

$P(X = x_2) = 0{,}1 + 0{,}25 + 0{,}2 = 0{,}55$,

отже, закон розподілу компоненти X:

X	x_1	x_2
P	0,45	0,55

Закон розподілу компоненти Y:

$P(Y = y_1) = 0{,}15 + 0{,}1 = 0{,}25$,

$P(Y = y_2) = 0{,}25 + 0{,}25 = 0{,}5$,

$P(Y = y_3) = 0{,}05 + 0{,}2 = 0{,}25$, отже,

Y	y_1	y_2	y_3
P	0,25	0,5	0,25

<u>Відповідь:</u> закон розподілу компоненти X:

X	x_1	x_2
P	0,45	0,55

закон розподілу компоненти Y:

Y	y_1	y_2	y_3
P	0,25	0,5	0,25

Для системи неперервних випадкових величин табличний спосіб задання закону розподілу ймовірностей неможливий. Тому, як і для одновимірних випадкових величин, закон розподілу ймовірностей неперервних випадкових величин задають функцією розподілу.

9.2. Функція розподілу системи двох випадкових величин

<u>*Означення.*</u> *Функцію розподілу системи двох випадкових величин* (X,Y) називається ймовірність сумісного виконання двох нерівностей $X < x$ і $Y < y$, тобто

$$F(x,y) = P(X < x; Y < y) \tag{9.1}$$

Зокрема для системи дискретних випадкових величин

$$F(x,y) = \sum_{x_i < x} \sum_{y_j < y} p_{ij}. \tag{9.2}$$

Функція розподілу системи випадкових величин називається також інтегральною функцією розподілу.

З геометричної точки зору значення функції розподілу $F(x,y)$ – це ймовірність влучення випадкової точки (X,Y) у нескінчений квадрат з вершиною у точці $M(x,y)$, розміщений нижче та лівіше від цієї вершини (рис. 9.1).

Права і верхня границі у квадраті не включаються.

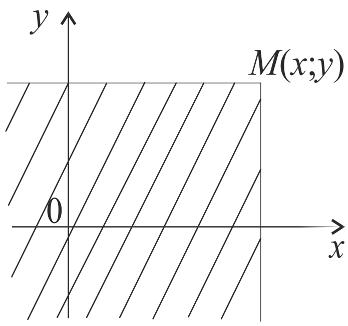
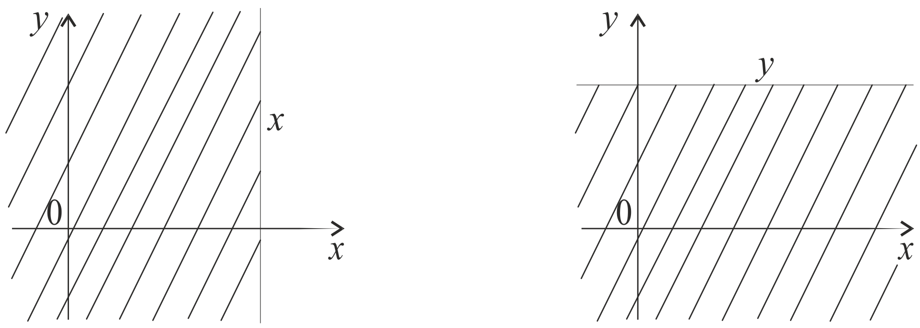


Рис. 9.1.

У цій інтерпретації функція розподілу однієї випадкової величини X – $F_1(x)$, представляє ймовірність влучення випадкової точки в праву частину півплощини, яка обмежена прямою, паралельною осі ординат і проходить через абсцису x (рис. 9.2 (а)), а функція розподілу однієї випадкової величини Y – $F_2(y)$ – ймовірність влучення у півплощину, яка обмежена зверху прямою паралельною осі абсцис і проходить через ординату y (рис. 9.2 (б)).



а) б)

Рис. 9.2.

Основні властивості функції розподілу системи двох випадкових величин.

1. $0 \le F(x,y) \le 1$ (оскільки $F(x,y)$ – ймовірність).

2. $F(x,y)$ – неспадна, тобто

$$F(x_2,y) \ge F(x_1,y)\text{, коли } x_2 > x_1,$$

$$F(x,y_2) \ge F(x,y_1)\text{, коли } y_2 > y_1,$$

(у разі збільшення будь-якого аргументу область квадрата збільшується, отже, ймовірність влучення в нього випадкової точки не може зменшуватись).

3. $F(x,-\infty)=F(-\infty,y)=F(-\infty;-\infty)=0$.

4. Нехай $F_1(x)$ і $F_2(y)$ – функції розподілу відповідно випадкових величин X і Y, а $F(x,y)$ – функції розподілу системи випадкових величин. Тоді

$$F(x,+\infty)=F_1(x),$$

$$F(-\infty,y)=F_2(y).$$

5. $F(-\infty,+\infty)=1$ (оскільки одночасне настання події $x<+\infty$ і $y<+\infty$ є вірогідною подією).

Для знаходження ймовірності влучення системи випадкових величин $\left(X,Y\right)$ у прямокутник $R=\left\{(x,y): x_1\le X\le x_2; y_1\le Y\le y_2\right\}$ слід розглянути рис. 9.3.

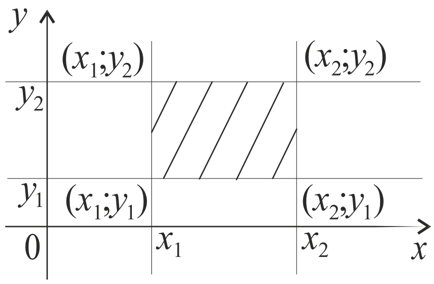


Рис. 9.3.

З рис. 9.3

$$P(x_1\le X\le x_2; y_1\le Y\le y_2)=F(x_2,y_2)-F(x_1,y_2)-F(x_2,y_1)+F(x_1,y_1).$$

9.3. Щільність розподілу ймовірностей двох випадкових величин

Нехай система двох неперервних випадкових величин $\left(X,Y\right)$ інтерпретується як точка координатної площини. На рис. 9.4 на координатній площині зображено малий прямокутник R_Δ зі сторонами Δx і Δy паралельними осям координат, одна із вершин якого має координати $(x;y)$.

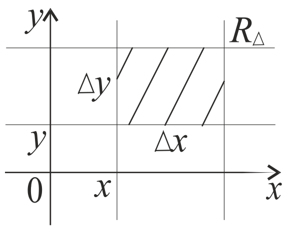


Рис. 9.4.

<u>*Означення.*</u> Функція $f(x,y)=\lim\limits_{\substack{\Delta x\to 0\\ \Delta y\to 0}}\dfrac{P\{(x,y)\in R_\Delta\}}{\Delta x\Delta y}$, якщо вона існує і неперервна, називається *щільністю розподілу ймовірностей системи випадкових величин*.

З означення щільності

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\partial y}.$$

Дійсно,

$$f(x,y)=\lim_{\substack{\Delta x\to 0\\ \Delta y\to 0}}\frac{F(x+\Delta x,y+\Delta y)-F(x+\Delta x,y)-F(x,y+\Delta y)+F(x,y)}{\Delta x\Delta y}=$$

$$=\lim_{\Delta x\to 0}\frac{1}{\Delta x}\left[\lim_{\Delta y\to 0}\frac{F(x+\Delta x,y+\Delta y)-F(x+\Delta x,y)}{\Delta y}-\lim_{\Delta y\to 0}\frac{F(x,y+\Delta y)-F(x,y)}{\Delta x\Delta y}\right]=$$

$$=\lim_{\Delta x\to 0}\frac{1}{\Delta x}\left[\frac{\partial F(x+\Delta x,y)}{\partial y}-\frac{\partial F(x,y)}{\partial y}\right]=\frac{\partial^2 F(x,y)}{\partial x\partial y}.$$

Основні властивості щільності розподілу ймовірностей системи двох випадкових величин.

1. Щільність розподілу двовимірної випадкової величини є невід'ємною функцією

$$f(x,y)\ge 0.$$

2. Інтегральна функція розподілу дорівнює

$$F(x,y)=\int_{-\infty}^{x}\int_{-\infty}^{y}f(t,z)dtdz.$$

3. Щільність розподілу складової X є

$$f_1(x)=\int_{-\infty}^{+\infty} f(x,z)dz,$$

щільність розподілу складової Y є

$$f_2(x)=\int_{-\infty}^{+\infty} f(t,y)dt.$$

4. Подвійний невласний інтеграл з нескінченими межами інтегрування від щільності розподілу $f(x,y)$ дорівнює одиниці

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)dxdy=1.$$

5. Ймовірність попадання двовимірної випадкової величини в область D знаходиться за формулою

$$P(X,Y)=\iint_D f(x,y)dxdy\,.$$

<u>Наслідок.</u> Ймовірність попадання двовимірної випадкової величини в прямокутник $R=\{(x,y):x_1\le X\le x_2;y_1\le Y\le y_2\}$ знаходиться за формулою

$$P(X,Y)=\int_{x_1}^{x_2}\int_{y_1}^{y_2} f(x,y)dxdy.$$

<u>Приклад 9.2.</u> Знайти ймовірність попадання випадкової величини (X,Y) у круг з радіусом $r=1$ і центром у початку координат, якщо її щільність задана формулою:

$$f(x,y)=\begin{cases}\dfrac{3}{8\pi}\left(2-\sqrt{x^2+y^2}\right), & x^2+y^2\le 1;\\ 0, & x^2+y^2>1.\end{cases}$$

<u>Розв’язання.</u>

Для знаходження ймовірності попадання випадкової величини (X,Y) в область $\{D:x^2+y^2\le 1\}$ потрібно обчислити подвійний інтеграл від щільності розподілу по цій області

$$P\{(X,Y)\in D\}=\frac{3}{8\pi}\iint_D\left(2-\sqrt{x^2+y^2}\right)dxdy.$$

Для обчислення цього інтегралу потрібно перейти до полярних координат, тобто

$$P\{(X,Y)\in D\}=\frac{3}{8\pi}\int_0^{\pi}d\varphi\int_0^1(2-\rho)\rho d\rho=\frac{3}{8\pi}2\pi\left(\rho^2-\frac{\rho^3}{3}\right)\Bigg|_0^1=\frac{3}{4}\left(1-\frac{1}{3}\right)=\frac{1}{2}.$$

<u>Відповідь:</u> ймовірність попадання випадкової величини (X,Y) у круг з радіусом $r=1$ і центром у початку координат дорівнює $\frac{1}{2}$.

9.4. Умовний закон розподілу

Відомо, що для залежних випадкових подій A і B справедлива рівність

$$P_A(B)=\frac{P(AB)}{P(A)}.$$

Аналогічна рівність має місце і для випадкових величин. Для того, щоб охарактеризувати залежність між складовими двовимірної випадкової величини потрібно ввести поняття умовного закону розподілу.

<u>Означення.</u> Умовним законом розподілу компоненти X з двовимірної випадкової величини (X,Y) називається її закон розподілу, знайдений при умові, що інша компонента Y прийняла певне значення y_j.

Нехай $P(x_i / y_j)$ – ймовірність того, що X набуде значення x_i при умові, що $Y=y_j$. Тоді умовним розподілом компоненти X при $Y=y_j$ називається сукупність ймовірностей $P(x_1 / y_j), P(x_2 / y_j), ..., P(x_m / y_j)$, які обчислені при припущенні, що подія $Y=y_j$ вже відбулася.

Якщо відомо закон розподілу двовимірної випадкової величини (X,Y), то умовні закони розподілу складової X обчислюються за формулами

$$P(x_i / y_j)=\frac{P(x_i y_j)}{P(y_j)}=\frac{p_{ij}}{P(y_j)}, \tag{9.3}$$

складової Y –

$$P(y_j / x_i) = \frac{P(x_i y_j)}{P(x_i)} = \frac{p_{ij}}{P(x_i)}. \qquad (9.4)$$

<u>**Приклад 9.3.**</u> Дискретну двовимірну випадкову величину задано таблицею

$X \backslash Y$	y_1	y_2	y_3
x_1	0,1	0,3	0,2
x_2	0,06	0,18	0,16

Знайти умовний закон розподілу складової Y за умови, що $X = x_1$.

<u>Розв'язання.</u>

Спочатку потрібно знайти умовні ймовірності $P(y_1 / x_1)$, $P(y_2 / x_1)$, $P(y_3 / x_1)$, які визначають шуканий умовний розподіл.

Враховуючи, що $P(x_1) = 0{,}1 + 0{,}3 + 0{,}2 = 0{,}6$, за формулами (9.4) обчислюються

$$P(y_1 / x_1) = \frac{0{,}1}{0{,}6} = \frac{1}{6},\ P(y_2 / x_1) = \frac{0{,}3}{0{,}6} = \frac{1}{2},\ P(y_3 / x_1) = \frac{0{,}2}{0{,}6} = \frac{1}{3}.$$

Потрібно перевірити чи дорівнює сума знайдених ймовірностей 1:

$$\frac{1}{6} + \frac{1}{2} + \frac{1}{3} = 1.$$

Отже, знайдено умовний закон розподілу компоненти Y за умови, що $X = x_1$.

<u>Відповідь:</u> умовний закон розподілу компоненти Y за умови, що $X = x_1$:

$$P(y_1 / x_1) = \frac{1}{6},\ P(y_2 / x_1) = \frac{1}{2},\ P(y_3 / x_1) = \frac{1}{3}.$$

9.5. Незалежні випадкові величини

Означення. Дві випадкові величини називаються *незалежними*, якщо закон розподілу однієї з них не залежить від того, яких можливих значень набуде інша.

Теорема. Для того, щоб випадкові величини були незалежними, необхідно і достатньо, щоб інтегральна функція розподілу $F(x, y)$ дорівнювала добутку функцій розподілу компонент, тобто

$$F(x, y) = F_1(x)F_2(y). \tag{9.5}$$

Зауваження. Рівність (9.5) вказує на незалежність компонент X і Y двовимірної величини $\left(X, Y\right)$.

9.6. Числові характеристики двовимірної випадкової величини

Крім числових характеристик окремих компонент двовимірної випадкової величини, розглядаються також числові характеристики, які характеризують двовимірну випадкову величину сукупно.

Означення. Кореляційним моментом двовимірної випадкової величини $\left(X, Y\right)$ називається величина

$$\mu_{xy} = M\left[(X - M(X))(Y - M(Y))\right]$$

або

$$\mu_{xy} = M(XY) - M(X)M(Y).$$

Для обчислення кореляційного моменту дискретної випадкової величини використовують формулу

$$\mu_{xy} = \sum_{i=1}^{m}\sum_{j=1}^{n}\left(x_i - M(X)\right)\left(y_j - M(Y)\right)P_{ij}.$$

З означення кореляційного моменту випливають наступні його властивості.

Властивості кореляційного моменту двовимірної випадкової величини.

1. Якщо X і Y – незалежні, то $\mu_{xy} = 0$.

Дійсно, для незалежних випадкових величин X і Y

$M(XY) = M(X)M(Y)$, тоді $\mu_{xy} = M(X)M(Y) - M(X)M(Y) = 0$.

Зауваження. Якщо $\mu_{xy} = 0$, то звідси не завжди випливає незалежність випадкових величин X і Y.

2. $|\mu_{xy}| \le \sigma(X)\sigma(Y)$, де $\sigma(X) = \sqrt{D(X)}$, $\sigma(Y) = \sqrt{D(Y)}$.

Означення. Коефіцієнтом кореляції двовимірної випадкової величини (X, Y) називається величина

$$r_{xy} = \frac{\mu_{xy}}{\sigma_x \sigma_y}.$$

Безпосередньо із означення випливає, що r_{xy} – безрозмірна величина, така, що $|r_{xy}| \le 1$.

Зауваження. Якщо $r_{xy} \ne 0$, то випадкові величини X і Y корельовано, а якщо $r_{xy} = 0$, то некорельовані.

Приклад 9.4. Задано закон розподілу двовимірної випадкової величини таблицею (X, Y):

$X \backslash Y$	$y_1 = -1$	$y_2 = 0$	$y_3 = 1$	$y_4 = 2$
$x_1 = 1$	0,1	0,25	0,3	0,15
$x_2 = 2$	0,1	0,05	0	0,05

Обчислити μ_{xy} і r_{xy}.

Розв’язання. Спочатку потрібно знайти закони розподілу компонент X і Y

:

X	1	2
P	0,8	0,2

Y	-1	0	1	2
P	0,2	0,3	0,3	0,2

Потрібно обчислити $M(X)$, $D(X)$, $\sigma(X)$, $M(Y)$, $D(Y)$, $\sigma(Y)$, $M(XY)$:

$M(X)=1\cdot 0,8+2\cdot 0,2=1,2;$

$D(X)=M(X^2)-M^2(X)=1^2\cdot 0,8+2^2\cdot 0,2-(1,2)^2=0,16;$

$\sigma(X)=\sqrt{D(X)}=\sqrt{0,16}=0,4;$

$M(Y)=(-1)\cdot 0,2+0\cdot 0,3+2\cdot 0,2=0,5;$

$D(Y)=M(Y^2)-M^2(Y)=(-1)^2\cdot 0,2+0^2\cdot 0,3+2^2\cdot 0,2-(0,5)^2=1,05;$

$\sigma(Y)=\sqrt{D(Y)}=\sqrt{1,05}\approx 1,025;$

$M(XY)=\sum_{i=1}^{m}\sum_{j=1}^{n}x_i y_j p_{ij}=1\cdot 0,1\cdot(-1)+1\cdot 0,25\cdot 0+1\cdot 0,3\cdot 1+1\cdot 0,15\cdot 2+$

$+2\cdot 0,1\cdot(-1)+2\cdot 0,05\cdot 0+2\cdot 0\cdot 1+2\cdot 0,05\cdot 2=0,5.$

Отже, $\mu_{xy}=M(XY)-M(X)M(Y)=0,5-1,2\cdot 0,5=-0,1;$

$$r_{xy}=\frac{\mu_{xy}}{\sigma_x\sigma_y}=\frac{-0,1}{0,4\cdot 1,05}=-0,244.$$

Відповідь: $\mu_{xy}=-0,1;\ r_{xy}=-0,244.$

Практичне заняття 9

9.1. Знайти закони розподілу складових двовимірної випадкової величини, яка задана законом розподілу

а)

$X \backslash Y$	y_1	y_2	y_3
x_1	0,1	0,2	0,1
x_2	0,15	0,25	0,2

б)

$X \backslash Y$	y_1	y_2	y_3
x_1	0,05	0,25	0,05
x_2	0,3	0,1	0,25

в)

$X \backslash Y$	y_1	y_2	y_3
x_1	0,3	0,3	0,1
x_2	0,05	0,15	0,1

г)

$X \backslash Y$	y_1	y_2	y_3
x_1	0,2	0,15	0,15
x_2	0,1	0,25	0,15

9.2. Керівництво підприємства вирішило після курсів підвищення кваліфікації екзаменувати співробітників економічного відділу, у якому працює 15 чоловік. У результаті 3 склали екзамен на «відмінно», 7 – на «добре», 4 – на «задовільно», решта – не склали і їх відправили на проходження курсів підвищення кваліфікації ще раз, але вже за свій рахунок. Нехай X – кількість співробітників, які склали екзамен на «відмінно», Y – кількість співробітників, які склали екзамен на «добре». Знайти: 1) закон розподілу двовимірної випадкової величини (X,Y); 2) закони розподілу складових X та Y.

9.3. Два студенти на великій перерві в університеті заходять в їдальню поїсти пиріжків. Кожний з'їдає від одного до трьох пиріжків. Закони розподілу кількості пиріжків, які з'їли хлопці представлені у таблицях:

I студент	1	2	3
p	0,2	0,7	0,1

II студент	1	2	3
$x_1 = 1$	0,1	0,45	0,45

Прибуток від продажу одного пиріжка складає 2 грн. Знайти закон розподілу прибутку, отриманого їдальнею від візиту цих двох студентів.

9.4. Задано закон розподілу двовимірної випадкової величини (X,Y) таблицею:

а)

$X \backslash Y$	$y_1 = -1$	$y_2 = 0$	$y_3 = 2$	$y_4 = 5$
$x_1 = -1$	0,15	0,15	0,2	0,14
$x_2 = 1$	0,06	0,05	0,1	0,15

б)

$X \backslash Y$	$y_1 = 0$	$y_2 = 1$	$y_3 = 4$	$y_4 = 7$
$x_1 = -2$	0,1	0	0,3	0,1
$x_2 = 0$	0,25	0,05	0,15	0,05

в)

Y \ X	$y_1 = -3$	$y_2 = -1$	$y_3 = 0$	$y_4 = 3$
$x_1 = -1$	0,05	0,25	0,1	0,15
$x_2 = 0$	0,1	0,05	0,05	0
$x_3 = 3$	0	0,1	0,05	0,1

г)

Y \ X	$y_1 = -5$	$y_2 = -3$
$x_1 = 1$	0,3	0,25
$x_2 = 2$	0,2	0,05
$x_3 = 4$	0,05	0,15

Знайти: 1) закони розподілу складових X та Y; 2) умовний закон розподілу складової Y за умови, що $X = x_1$; 3) умовний закон розподілу складової X за умови, що $Y = y_2$; 4) ймовірність події $\{X < x_2, Y > y_2\}$.

9.5. Знайти значення сталої C та ймовірність події $\{x + y < 2\}$, якщо щільність ймовірностей двовимірної випадкової величини (X, Y) задана формулою:

$$f(x, y) = \begin{cases} C(xy + y)^2, 0 \le x \le 1, 0 \le y \le 1; \\ 0, \text{у інших випадках.} \end{cases}$$

9.6. Двовимірна випадкова величина (X, Y) розподілена у квадраті $0 \le x \le 1$, $0 \le y \le 1$ з щільність ймовірностей $f(x, y) = 3xy(2 - y)$. Знайти ймовірність подій: 1) $X > Y$; 2) $Y - X > 0,5$; 3) $Y > 0,5$.

9.7. Задано закон розподілу двовимірної випадкової величини (X, Y) таблицею:

а)

$X \backslash Y$	$y_1=-1$	$y_2=0$	$y_3=2$	$y_4=5$
$x_1=-2$	0,14	0,15	0,2	0,15
$x_2=1$	0,06	0,05	0,1	0,15

б)

$X \backslash Y$	$y_1=0$	$y_2=1$	$y_3=3$	$y_4=7$
$x_1=-7$	0,15	0	0,25	0,15
$x_2=1$	0,2	0,05	0,17	0,03

в)

$X \backslash Y$	$y_1=-2$	$y_2=-1$	$y_3=1$	$y_4=3$
$x_1=-3$	0	0,05	0,05	0,1
$x_2=0$	0,05	0,25	0,1	0
$x_3=5$	0,15	0,1	0,05	0,1

г)

$X \backslash Y$	$y_1=-3$	$y_2=-1$	$y_3=0$
$x_1=-1$	0,1	0,15	0
$x_2=1$	0,2	0	0,05
$x_3=4$	0,3	0,15	0,05

Обчислити μ_{xy} і r_{xy}.

Зразки самостійних та контрольних робіт

Самостійна робота
на тему «Формули обчислення ймовірностей. Повторні незалежні випробування»

Варіант 1

1. На концерт виділено 10 мікрофонів, причому 3 із них бездротові. Знайти ймовірність того, що серед 4 навмання взятих мікрофонів, 2 виявляться бездротовими.

2. Документацію, яка поступає на підприємство, розподіляють між відділами 3 кур’єри. Ймовірності того, що документація попаде не за адресою, при розподілі першим кур’єром 0,2; другим – 0,3; третім – 0,1. Знайти ймовірність того, що хоча б один кур’єр допустить помилку при розподілі документації.

3. На складі міститься однорідна продукція трьох заводів: 120 ящиків заводу № 1, 130 – № 2, 150 – № 3. Ймовірність того, що продукція відповідає стандарту для кожного заводу відповідно становить 0,9; 0,95; 0,75. Навмання взятий зі складу ящик виявився стандартним. Яка ймовірність того, що це продукція заводу № 2?

4. З практики відомо, що 3 % річних звітів підприємств при аудиторській перевірці повертається на доопрацювання. Яка ймовірність того, що із навмання перевірених 80 звітів рівно 3 будуть повернені, на доопрацювання?

Варіант 2

1. У фірмі працюють 7 чоловіків і 5 жінок. За табельними номерами відібрали 3 особи. Яка ймовірність, що серед них 3 жінки?

2. Зроблено три постріли по мішені. Ймовірність влучення при першому пострілі дорівнює 0,7; при другому – 0,85; при третьому – 0,95. Яка ймовірність того, що буде лише одне влучення?

3. Сировина на підприємство надходить від 2 постачальників, у кількісному співвідношенні 1:3. Високоякісна сировина у цих постачальників становить відповідно 85 % і 80 %. Вибрана навмання сировина виявилася високоякісною. Від якого постачальника вона найімовірніше надійшла?

4. Завод відправив у торгівельну мережу партію із 10000 одиниць продукції. Частка пошкоджених при транспортуванні виробів становить 0,01 %. Знайти ймовірність того, що при транспортуванні буде пошкоджено 2 вироби.

Залікова контрольна робота
з теорії ймовірностей

Варіант 1

1. Ймовірність того, що деталь стандартна дорівнює 0,7. Знайти ймовірність того, що з трьох перевірених деталей тільки одна стандартна.

2. На конвеєр надходять пляшки із водою з трьох автоматизованих ліній. Перша має 0,2 % браку, друга – 0,25 %, третя – 0,3 %. Знайти ймовірність надходження на конвеєр бракованої пляшки, якщо від кожного автомата надійшло відповідно 100, 150 та 250 пляшок.

3. В середньому 15 % пакетів акцій на аукціонах продаються за початково-заявленою ціною. Знайти ймовірність того, що із 7 виставлених на торгах пакетів акцій, принаймні 3 пакети будуть продані за початково-заявленою ціною.

4. Задано закон розподілу дискретної випадкової величини:

X	1	2	3
P	0,4	0,3	0,3

Знайти $M(X)$, $D(X)$, функцію розподілу $F(x)$, нарисувати її графік.

5. Неперервна випадкова величина задана функцією розподілу:

$$F(x)=\begin{cases}0, & \text{при } x\le 2;\\ (2x-4)^2, & \text{при } 2<x\le \dfrac{5}{2};\\ 1, & \text{при } x>\dfrac{5}{2}.\end{cases}$$

Знайти $P\left(2<x<\dfrac{7}{3}\right)$, $f(x)$.

6. Неперервна випадкова величина задана щільністю розподілу:

$$f(x)=\begin{cases}0, & \text{при } x\le 0;\\ \frac{1}{2}x, & \text{при } 0<x\le 2;\\ 0, & \text{при } x>2.\end{cases}$$

Знайти $M(X)$, $D(X)$, функцію розподілу $F(x)$.

Варіант 2

1. Фірмою відряджений автомобіль за матеріалами на 4 бази. Ймовірність наявності потрібного матеріалу на першій базі дорівнює 0,6, на другій – 0,7, на третій – 0,75, на четвертій – 0,5. Знайти ймовірність того, що принаймні на трьох базах будуть необхідні матеріали.

2. Продукція цеху перевіряється двома контролерами, причому перший контролер перевіряє 65 % виробів, а другий – решту. Ймовірність того, що перший контролер пропустить неякісний виріб, дорівнює 0,02, другий – 0,03. Взятий навмання виріб, маркований як якісний, виявився неякісним. Знайти ймовірність того, що цей виріб перевіряв другий контролер.

3. Завод відправив у торгівельну мережу партію із 10000 одиниць продукції. Частка пошкоджених при транспортуванні виробів становить 0,02 %. Знайти ймовірність того, що при транспортуванні буде пошкоджено 3 вироби.

4. Задано закон розподілу дискретної випадкової величини:

X	-1	2	3
P	0,5	0,3	0,2

Знайти $M(X)$, $D(X)$, функцію розподілу $F(x)$, нарисувати її графік.

5. Неперервна випадкова величина задана функцією розподілу:

$$F(x)=\begin{cases}0, & \text{при } x\le 0;\\ x^3, & \text{при } 0<x\le 1;\\ 1, & \text{при } x>1.\end{cases}$$

Знайти $P\left(\frac{1}{2}<x<1\right)$, $f(x)$.

6. Неперервна випадкова величина задана щільністю розподілу:

$$f(x)=\begin{cases}0, & \text{при } x\le \frac{1}{2};\\ 8(4x-2), & \text{при } \frac{1}{2}<x\le \frac{3}{4};\\ 0, & \text{при } x>\frac{3}{4}.\end{cases}$$

Знайти $M(X)$, $D(X)$, функцію розподілу $F(x)$.

ПРЕДМЕТНИЙ ПОКАЖЧИК

Частина II. Математична статистика

Лекція 1. Вибірки та їх представлення

1.1. Предмет і методи математичної статистики

1.2. Генеральна і вибіркові сукупності

1.3. Статистичний розподіл вибірки

1.4. Емпірична функція розподілу

1.5. Графічне зображення статистичних розподілів

1.1. Предмет і методи математичної статистики

Математична статистика займається встановленням закономірностей масових явищ і процесів на основі результатів спостережень або експериментів над ними. Для виявлення таких закономірностей використовуються математичні методи:

– збирання і систематизація статистичних даних;

– наукового аналізу даних.

Математична статистика опирається на теорію ймовірностей. Застосування математичної статистики в економіці дозволяє будувати економічні моделі і оцінювати їхні параметри, перевіряти твердження (гіпотези) про властивості економічних показників та форми їх взаємозв'язку. Одержані результати є основою для економічного аналізу і прогнозування та дають можливість приймати обґрунтовані економічні рішення у бізнесі в умовах невизначеності.

1.2. Генеральна і вибіркові сукупності

Нехай потрібно дослідити сукупність однорідних об'єктів відносно деякої якісної або кількісної ознаки. Якісні ознаки характеризують деякі властивості або стан об'єкта, наприклад, стать, професія, якість продукції тощо. Кількісні ознаки одержують у результаті вимірювання спостереження у вигляді числових значень, наприклад, маса, об’єм, прибуток тощо. Кількісні ознаки позначаються через X, Y, Z вони можуть бути дискретними (ізольовані значення) або неперервними, тобто набувати будь-яких числових значень у деяких межах.

Можливі два способи дослідження ознаки сукупності: суцільне спостереження і вибіркове спостереження. При суцільному спостереженні вивчається кожний об’єкт сукупності, а при вибірковому деякої частини, відібраної із усієї сукупності. При цьому вся сукупність об’єктів, яка підлягає дослідженню, називається *генеральною*, а та її частина, яка попала на перевірку або дослідження, *вибірковою сукупністю* або *вибіркою*. Число об’єктів у генеральній сукупності і у вибірці називається їх о*бсягами*. За результатами вивчення ознаки у вибірці робиться висновок про властивості генеральної сукупності.

Дослідження вибіркової сукупності дає такі переваги:

1) практичність (із-за великого обсягу неможливо охопити всю генеральну сукупність);

2) виграш у часі;

3) зменшення затрат праці;

4) достатньо високу достовірність;

5) економію продукції.

Інформація про генеральну сукупність, одержана на основі вибірки, завжди буде мати деяку похибку, оскільки ґрунтується на вивченні тільки частини об’єктів. Тому дуже важливо так організувати вибіркове спостереження, щоб згадана інформація була найбільш повною і давала правильне уявлення про генеральну сукупність. Неправильно сформована вибірка дасть викривлене уявлення про сукупність.

Розрізняють два способи відбору елементів генеральної сукупності у вибірку: випадковий і невипадковий. Випадковий відбір означає, що кожен об’єкт сукупності має рівний шанс потрапити до вибірки. Якщо сукупність містить групу об'єктів, то випадковий відбір повинен забезпечити представництво у вибірці об'єктів кожної із цих груп. Такий відбір називають типовим. Наприклад, продукція виробляється різними цехами, на різних конвеєрах. Тому при вивченні якості виробів підприємства доцільно робити вибірку із продукції кожного цеху і кожного конвеєра. Так, при наявності трьох цехів і двох ліній у кожному, продукцію доцільно розбити на шість груп.

Випадковий відбір, як правило, проводять за допомогою спеціальних таблиць випадкових чисел. Для цього елементи сукупності нумеруються і з таблиці випадкових чисел, відкритої навмання на довільній сторінці, виписуються підряд номера елементів, які повинні ввійти у вибірку. Останнім часом використовують комп'ютер, який генерує випадкові числа.

Серед невипадкових вибірок виділяють два види вибірок: серійний і механічний. Серійним називають відбір, при якому об'єкти вибираються із генеральної сукупності не по одному, а серіями, на які попередньо поділена сукупність. При механічному відборі елементи генеральної сукупності відбираються за наперед встановленим правилом. Наприклад, до 5%-вої вибірки входить кожен двадцятий член генеральної сукупності.

Зауважимо, що при дослідженні таких вибірок не завжди можна застосовувати ймовірнісні методи.

1.3. Статистичний розподіл вибірки

Нехай X – випадкова величина, яка уособлює певну кількісну ознаку об'єктів генеральної сукупності. Із цієї сукупності одержано вибірку ознаки X : $X_1, X_2, \ldots, X_k$, де X_i, $i=\overline{1,k}$ – незалежні однаково розподілені випадкові величини, розподіл кожної з яких співпадає з розподілом X. Різні значення, які приймає кількісна ознака в наслідок відбору, наприклад, $X_i = x_i \quad (i=\overline{1,k})$ називаються *варіантами або реалізаціями*.

Послідовність варіант, розміщених у зростаючому порядку $x_1 < x_2 < ... < x_k$ називається *варіаційним рядом*. Якщо варіанта x_i зустрічається n_i разів у вибірці, то n_i називається *частотою* варіанти x_i $(i = \overline{1,k})$. При цьому $\sum_{i=1}^{k} n_i = n$ – обсяг вибірки. Відношення $\frac{n_i}{n} = w_i$ називається *відносною частотою* варіанти, причому $\sum_{i=1}^{k} w_i = 1.$ У залежності від того, яких значень можуть набувати варіанти, варіаційні ряди поділяються на *дискретні* та *неперервні* (*інтервальні*).

Означення. Перелік варіант варіаційного ряду і відповідних їм частот або відносних частот називають *статистичним розподілом вибірки:*

x_i	x_1	x_2	...	x_k
n_i	n_1	n_2	...	n_k

x_i	x_1	x_2	...	x_k
w_i	w_1	w_2	...	w_k

Приклад 1.1. На підприємстві по виробництву дитячого харчування досліджувався обсяг випуску яблучного пюре (т) протягом останніх десяти років.

Роки	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014
Обсяг (т)	4	3	4	6	6	8	7	6	8	7

За даними задачі побудувати статистичний розподіл вибірки.

Розв’язання. Досліджувана ознака X – обсяг випуску яблучного пюре (т). Варіаційний ряд матиме вигляд: $x_1 = 3,\quad x_2 = 4,\quad x_3 = 6,\quad x_4 = 7,\quad x_5 = 8.$ Статистичний розподіл частот:

x_i	3	4	6	7	8
n_i	1	2	3	2	2

$n = \sum_{i=1}^{5} n_i = 1 + 2 + 3 + 2 + 2 = 10;$

та відносних частот:

x_i	3	4	6	7	8
w_i	$0{,}1$	$0{,}2$	$0{,}3$	$0{,}2$	$0{,}2$

$n = \sum_{i=1}^{5} w_i = 1.$

Відповідь: статистичний розподіл частот:

x_i	3	4	6	7	8
n_i	1	2	3	2	2

відносних частот:

x_i	3	4	6	7	8
w_i	$0{,}1$	$0{,}2$	$0{,}3$	$0{,}2$	$0{,}2$

Для неперервного варіаційного ряду статистичний розподіл задається таблицею

інтервал	$[\alpha_0;\alpha_1)$	$[\alpha_1;\alpha_2)$	...	$[\alpha_{k-1};\alpha_k)$
n_i	n_1	n_2	...	n_k
w_i	w_1	w_2	...	w_k

Тут n_i, w_i – частота і відносна частота появи ознаки X в інтервалі $[\alpha_{i-1};\alpha_i)$. Довжина $h_i = \alpha_i - \alpha_{i-1}$ цього інтервалу називається *кроком*.

При інтервальному статистичному розподілі доцільно використовувати поняття щільності частоти $f_i = \dfrac{n_i}{h_i}$ або щільності відносної частоти $g_i = \dfrac{w_i}{h_i}$, які характеризують число елементів або відносне їх число, яке припадає на одиницю довжини h_i інтервалу.

Для зручності розрахунків, за звичай, крок обирається однаковим для даної вибірки. На практиці кількість інтервалів визначається за формулами: $k = [1 + 3{,}322\lg n]$ (Стерджес), $k = [5\ln n]$ (Брукс-Карузерс) або $k = \left[\sqrt{n}\right]$, де $[a]$ – ціла частина числа a. Крок інтервалу h_i визначається за формулою: $h_i = \dfrac{x_{\max} - x_{\min}}{k}$, де $x_{\max}$ – варіанта з найбільшим значенням, $x_{\min}$ – варіанта з найменшим значенням, k – кількість інтервалів.

Шляхом заміни $x_i = \dfrac{\alpha_i + \alpha_{i-1}}{2}$, $\left(i = \overline{1,k}\right)$ можна перейти від інтервального статистичного розподілу до дискретного.

1.4. Емпірична функція розподілу

Нехай досліджується кількісна ознака X генеральної сукупності.

Накопичена частота дорівнює числу елементів сукупності, які мають значення ознаки X менше за задане число x $(X < x)$, позначається через $H_n(x)$.

Кумулятивна частота (*відносна частота*) дорівнює відносному числу елементів, які мають значення ознаки X менше за задане число x $(X < x)$, позначається через $F_n(x)$. Функція $F_n(x)$ ще називається *емпіричною функцією розподілу*.

Отже, *емпірична функція розподілу* $F_n(x)$ для значення x визначає відносну накопичену частоту події $X < x$:

$$F_n(x) = W(X < x) = \frac{\mu(x)}{n},$$

де $\mu(x)$ – накопичена частота, тобто кількість варіант вибірки, менших за варіанту x.

Графік функції $F_n(x)$ називається *кумулятою* і є східчастою лінією, яка має точки розриву (скінчені стрибки) у точках $x_1,\ x_2,\ \ldots$.

На відміну від емпіричної функції розподілу вибірки, функція розподілу $F(x)$ генеральної сукупності називається *теоретичною функцією розподілу*. Відмінність між цими функціями полягає в тому, що $F(x)$ визначає ймовірність події $X < x$, а $F_n(x)$ – відносну частоту цієї ж події. При великих значеннях n $F_n(x)$ мало відрізняється від $F(x)$.

Для знаходження та побудови функції розподілу в створюється модуль, у якому використовуються функції та оператори описані у Ч. І (с. 72 – 73).

Потрібно звернути увагу на те, що у разі, якщо вибірка даних міститься у файлі програми *MS Excel*, то є декілька способів імпортування даних у програму *MathCad*, з ними можна ознайомитись у додатку 10.

Приклад 1.2. Для даного статистичного розподілу вибірки побудувати емпіричну функцію розподілу:

x_i	1	4	6	9
n_i	5	10	3	2

Розв’язання. Обсяг вибірки $n = 5 + 10 + 3 + 2 = 20.$

1. Оскільки x_1 – найменша із варіант, то при $x \le 1$ – $\mu(x) = 0$ і $F_n(x) = 0$.

2. Значення випадкової величини $X < 4$, зокрема $x_1 = 1$, спостерігалось 5 разів, тому $\mu(x) = 5$ і $F_n(x) = \frac{5}{20} = 0,25$ при $1 < x \le 4$.

3. Якщо значення випадкової величини знаходиться в інтервалі $4 < x \le 6$, то нерівність $X < x$ виконується для варіант $x_1 = 1, x_2 = 4$ і $\mu(x) = 5 + 10 = 15$, звідки $F_n(x) = \frac{15}{20} = 0,75$.

4. Для значень випадкової величини з інтервалу $6 < x \le 9$, нерівність $X < x$ виконується для варіант $x_1 = 1, x_2 = 4, x_3 = 6$ і $\mu(x) = 5 + 10 + 3 = 18$, звідки $F_n(x) = \frac{18}{20} = 0,9$.

5. Оскільки $x_4 = 9$ – найбільша варіанта, то $F_n(x) = \frac{20}{20} = 1$ при всіх значеннях випадкової величини $X > 9$.

Отже, емпірична функція розподілу матиме вигляд

$$F_n(x) = \begin{cases} 0 & \text{при } x \le 1; \\ 0,25 & \text{при } 1 < x \le 4; \\ 0,75 & \text{при } 4 < x \le 6; \\ 0,9 & \text{при } 6 < x \le 9; \\ 1 & \text{при } x > 9 \end{cases}$$

та її графік (рис. 1.1).

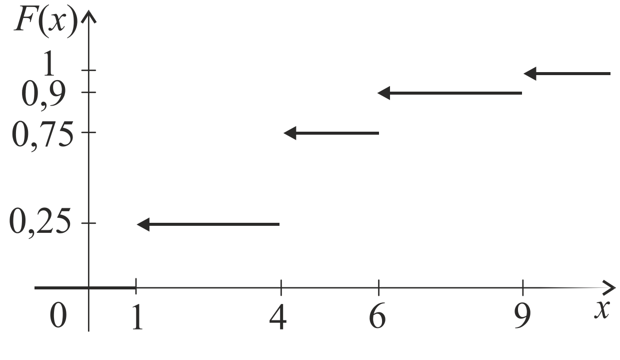


Рис. 1.1

$ORIGIN := 1$

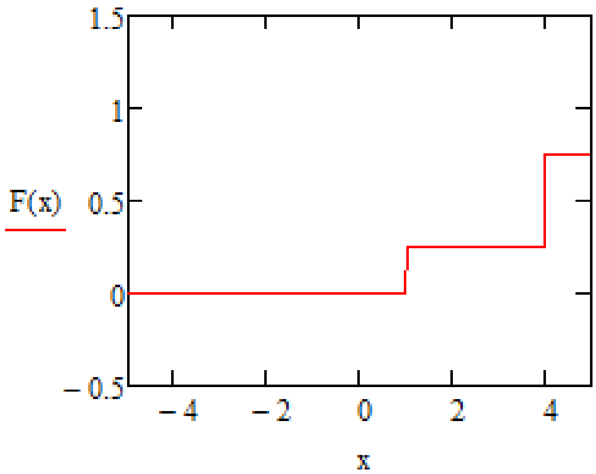
$X := (1 \;\; 4 \;\; 6 \;\; 9) \qquad N := (5 \;\; 10 \;\; 3 \;\; 2)$

$n := \sum N = 20$

$$P := \left| \begin{array}{l} \text{for } i \in 1..\text{cols}(N) \\ \quad P_{1,i} \leftarrow \dfrac{N_{1,i}}{n} \\ P \end{array} \right.$$

$P = (0.25 \;\; 0.5 \;\; 0.15 \;\; 0.1)$

$$F(x) := \left| \begin{array}{l} p \leftarrow 0 \\ i \leftarrow 1 \\ \text{while } x > X_{1,i} \\ \quad \left| \begin{array}{l} p \leftarrow p + P_{1,i} \\ i \leftarrow i + 1 \\ \text{break if } i > \text{cols}(X) \end{array} \right. \\ p \end{array} \right.$$



Відповідь: $F_n(x)=\begin{cases}0 & \text{при } x\le 1;\\ 0,25 & \text{при } 1<x\le 4;\\ 0,75 & \text{при } 4<x\le 6;\\ 0,9 & \text{при } 6<x\le 9;\\ 1 & \text{при } x>9.\end{cases}$

1.5. Графічне зображення статистичних розподілів

Графічна ілюстрація статистичних даних надає їм наглядності і часто приводить до спрощення їх аналізу.

У залежності від вигляду варіаційного ряду і поставленої задачі можна отримати різні графіки статистичних розподілів, зокрема, *полігон, гістограму, кумулятивну криву.*

Розглянемо дискретний статистичний розподіл вибірки. *Полігоном частот* (або *відносних частот*) називається ламана лінія, відрізки якої з'єднують точки з координатами (x_i, n_i) (або (x_i, w_i)), $i=\overline{1,k}$.

У випадку неперервного статистичного розподілу *гістограмою частот* (або *відносних частот*) називається східчаста фігура у вигляді послідовності прямокутників з основами $h_i=\alpha_i-\alpha_{i-1}$ і висотами f_i (або g_i). Площа S_i і-го прямокутника $S_i=h_i\cdot f_i=n_i$. Звідси, площа гістограми частот дорівнює обсягу вибірки: $\sum_{i=1}^{k}S_i=n.$ Аналогічно, для гістограми відносних частот: $S_i=h_i\cdot g_i=w_i$ і $\sum_{i=1}^{k}S_i=1.$

Розподіл ознаки у варіаційному ряді за накопиченими частотами або відносними частотами зображується у вигляді *кумуляти* або *кумулятивної кривої*. Для її побудови на осі абсцис відкладається значення ознаки, а на осі ординат – накопичені частоти або накопичені відносні частоти.

Для інтервального варіаційного ряду емпірична функція розподілу співпадає з кумулятою, побудованою за накопиченими відносними частотами, якщо її довизначити.

Для побудови гістограми частот:

необхідно вибрати пункт **Данные/Анализ данных** і у запропонованому переліку вибрати **Гистограмма**. У діалоговому вікні потрібно задати наступні параметри: *Входной інтервал* – вводяться посилання на клітинку або діапазон, які містять вибіркові дані; *Интервал карманов* (необов'язковий параметр) – вводяться посилання на клітинку або діапазон, які містять набір граничних значень для визначення інтервалів (*карманов*). Ці значення повинні бути введені у порядку зростання. Програма *MS Excel* обчислює кількість попадань даних у сформовані інтервали, причому межі інтервалів є строгими зліва і нестрогими справа. Якщо діапазон інтервалів не ввести, то набір інтервалів створиться автоматично (рівномірно розподілиться між мінімальними і максимальними значеннями даних); *Метки* – встановлюється в активний режим, якщо перший рядок вхідних даних містить заголовки; *Выходной интервал* – при встановленні в активний режим, вимагає ввода адреси верхньої клітинки, починаючи з якої будуть розміщені обчислені автоматично відносні частоти / *Новый рабочий лист / Новая рабочая книга* – встановлюється в активний режим для відкриття нового аркуша / нової книги, у яких будуть розміщені обчислені автоматично відносні; *Парето (отсортированная гистограмма)* – встановлюється в активний режим для представлення даних у порядку спадання частоти, інакше – дані будуть приведені у порядку слідування інтервалів; *Интегральный* *процент* – встановлюється в активний режим для розрахунку виражених у відсотках накопичених частот (*накопленных частостей*) і включення у гістограму графіка кумуляти; *Вывод графика* – встановлюється в активний режим для автоматичного створення вбудованої діаграми там, де міститься вихідний діапазон. Для побудови гістограми за допомогою описаного шляху потрібно мати початкову вибірку даних, якщо дані задано у вигляді інтервального статистичного розподілу необхідно вибрати пункт **Вставка/Гистограмма**. У діалоговому вікні задаються вхідні параметри;

необхідно призначити двовимірному графіку тип гістограми, для цього в діалоговому вікні **Formatting Currently Selected Graph** (Форматирование выбранного графика), яке визивається вибором правої клавіші миші (коли курсор знаходиться на графіку), у вкладці **Traces** (Трассировка) потрібно встановити тип (**Type**) списку **bar** (столбики) або **solidbar** (сплошные столбики). Функції та оператори, які використовуються для створення модуля описані у Ч. І (с. 72 – 73). Функції **rows(*A*)**, **cols(*A*)** описані у ч. І (с. 63).

Приклад 1.3. Для дослідження рівня знань з економічної теорії серед студентів курсу випадковим чином було відібрано та перевірено одну з груп. Було отримано наступні результати:

Бали	[30;45)	[45;60)	[60;75)	[75;90)	[90;100]
Кількість студентів	1	3	9	8	4

За даними перевірки побудувати: 1) гістограму відносних частот; 2) полігон відносних частот; 3) емпіричну функцію розподілу та її графік.

<u>Розв'язання.</u> Кількісна ознака X – середній бал з економічної теорії, має неперервний розподіл.

1) Знайшовши відносні частоти $w_i = \dfrac{n_i}{n}$ появи ознака X на відповідних інтервалах, можна отримати ряд розподілу:

Інтервал $[\alpha_{i-1};\alpha_i)$	$[30;45)$	$[45;60)$	$[60;75)$	$[75;90)$	$[90;100]$
w_i	$0,04$	$0,12$	$0,36$	$0,32$	$0,16$

Довжина інтервалів $h_i = \alpha_i - \alpha_{i-1} = 15,\ i = \overline{1,4},\ h_5 = 10$. Щільність відносної частоти g_i:

$$g_1 = \frac{0,04}{15} \approx 0,003, \quad g_2 = \frac{0,12}{15} \approx 0,008, \quad g_3 = \frac{0,36}{15} \approx 0,024, \quad g_4 = \frac{0,32}{15} \approx 0,021,$$

$$g_5 = \frac{0,16}{15} \approx 0,011.$$

Гістограма відносних частот має вигляд (рис. 1.2):

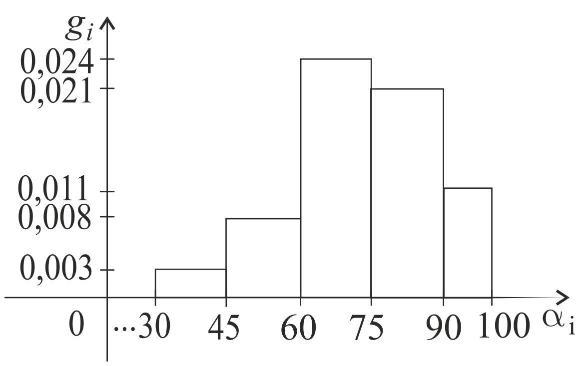


Рис. 1.2

2) Для побудови полігону відносних частот потрібно перейти від інтервального до дискретного статистичного розподілу відносних частот, шляхом заміни $x_i = \frac{\alpha_{i-1}+\alpha_i}{2}$, $\left(i=\overline{1,5}\right)$.

x_i	37,5	52,5	67,5	82,5	95
w_i	0,04	0,12	0,36	0,32	0,16

Полігон відносних частот зображено на рис. 1.3:

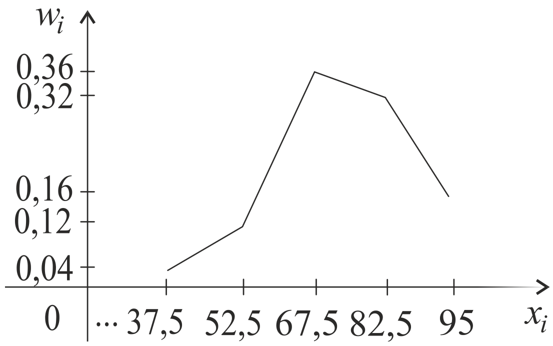


Рис. 1.3

3) а) Якщо $x \in [0;30]$, то $F_n(x) = \frac{\mu(x)}{n} = 0$, оскільки $\mu(x) = 0$ – кількість студентів, що отримали середній бал менше x, дорівнює нулю.

б) Якщо $x = 45$, то $F_n(x) = \frac{n_1}{n} = \frac{1}{25} = 0{,}04.$

в) Якщо $x = 60$, то $F_n(x) = \frac{n_1+n_2}{n} = \frac{1+3}{25} = 0{,}16.$

г) Якщо $x = 75$, то $F_n(x) = \frac{n_1+n_2+n_3}{n} = \frac{1+3+9}{25} = 0{,}52.$

д) Якщо $x = 90$, то $F_n(x) = \frac{n_1+n_2+n_3+n_4}{n} = \frac{1+3+9+8}{25} = 0{,}84.$

е) Якщо $x \geq 100$, то $F_n(x) = \frac{n_1+n_2+n_3+n_4+n_5}{n} = \frac{1+3+9+8+4}{25} = 1.$

Отже, емпірична функція розподілу має вигляд:

$$F_n(x)=\begin{cases}0 & \text{при } x\in[0;30];\\ 0,04 & \text{при } x=45;\\ 0,16 & \text{при } x=60;\\ 0,52 & \text{при } x=75;\\ 0,84 & \text{при } x=90;\\ 1 & \text{при } x\ge 100.\end{cases}$$

Її графік має вигляд довизначеної кумулятивної кривої (рис. 1.4):

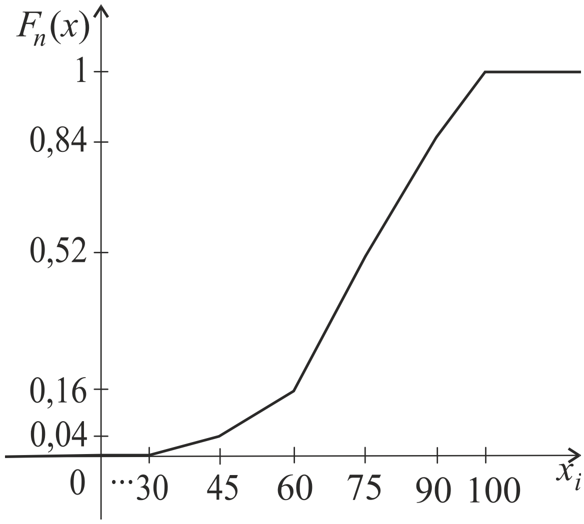


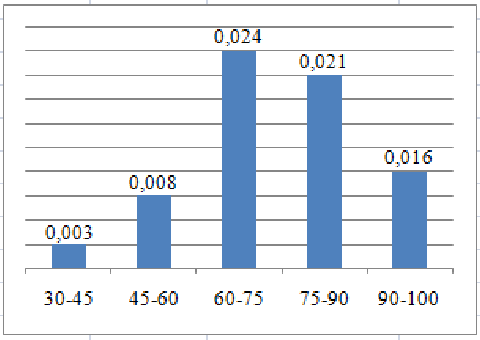
Рис. 1.4

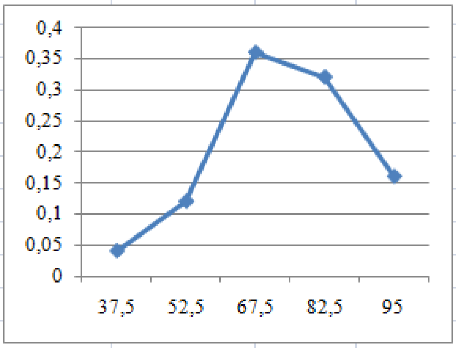
	A	B	C	D	E	F	G	H
1	$[\alpha_{i-1};\alpha_i)$	Довжина інтервала	$x_i=\dfrac{\alpha_{i-1}+\alpha_i}{2}$	Частота	Відносна частота	Щільність відн.частоти		
2	30-45	15	37,5	1	0,04	0,00266667		
3	45-60	15	52,5	3	0,12	0,008		
4	60-75	15	67,5	9	0,36	0,024		
5	75-90	15	82,5	8	0,32	0,02133333		
6	90-100	10	95	4	0,16	0,016		
7	**обсяг**			**25**				

=СУММ(D2:D6)

=D6/D7

=E6/B6



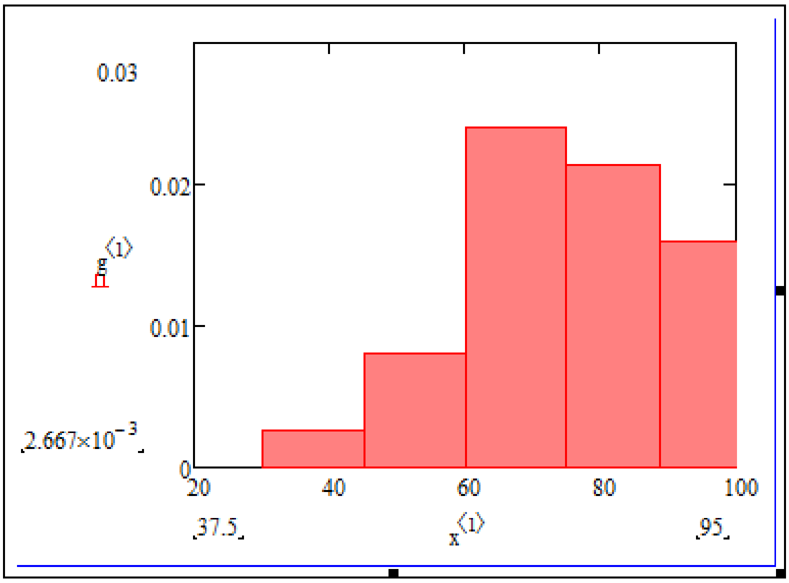


$ORIGIN := 1$

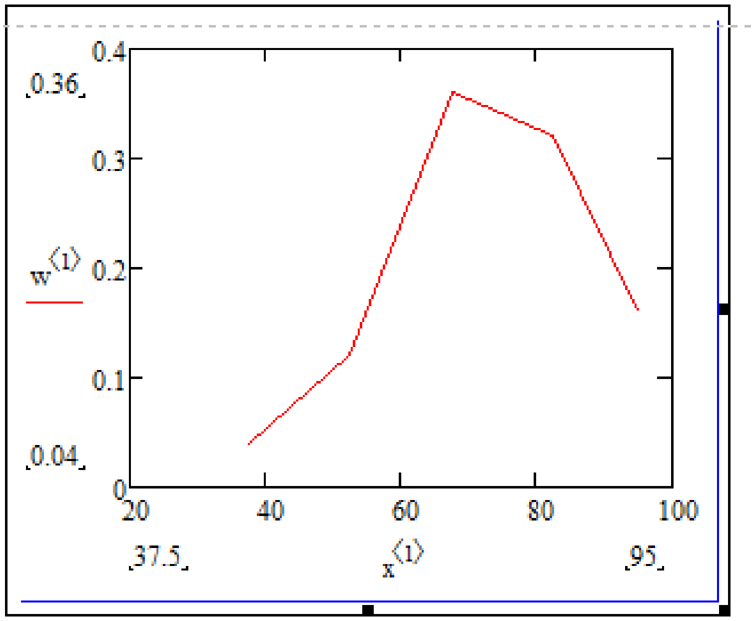
$$IN := \begin{pmatrix} 30 & 45 & 1 \\ 45 & 60 & 3 \\ 60 & 75 & 9 \\ 75 & 90 & 8 \\ 90 & 100 & 4 \end{pmatrix} \qquad n := \sum IN^{\langle 3 \rangle} = 25$$

$$g := \left| \begin{array}{l} \text{for } i \in 1..\,\text{rows}(IN) \\ \quad g_{i,1} \leftarrow \dfrac{IN_{i,3}}{n \cdot \left(IN_{i,2} - IN_{i,1}\right)} \\ g \end{array} \right. \qquad g = \begin{pmatrix} 2.667 \times 10^{-3} \\ 8 \times 10^{-3} \\ 0.024 \\ 0.021 \\ 0.016 \end{pmatrix}$$

$$x := \left| \begin{array}{l} \text{for } i \in 1..\,\text{rows}(IN) \\ \quad x_{i,1} \leftarrow \dfrac{IN_{i,1} + IN_{i,2}}{2} \\ x \end{array} \right. \qquad x = \begin{pmatrix} 37.5 \\ 52.5 \\ 67.5 \\ 82.5 \\ 95 \end{pmatrix}$$



$$w := \left|\begin{array}{l} \text{for } i \in 1..\text{rows}(IN) \\ \quad w_{i,1} \leftarrow \dfrac{IN_{i,3}}{n} \\ w \end{array}\right. \qquad w = \begin{pmatrix} 0.04 \\ 0.12 \\ 0.36 \\ 0.32 \\ 0.16 \end{pmatrix}$$



Практичне заняття 1

1.1. Підприємство, яке спеціалізується на оптовій реалізації овочів та фруктів, вирішило дослідити середньомісячну потребу у картоплі (у кг) серед основних споживачів. Дослідження дало результати: 200, 150, 250, 100, 150, 200, 200, 300, 500, 300. Скласти дискретний статистичний розподіл частот та відносних частот.

1.2. Під час дослідження кількісної ознаки *X* із генеральної сукупності було отримано вибірку 5, 4, 8, 4, 7, 3, 4, 1, 2, 7, 4, 6, 6, 4, 7, 3, 5, 1, 2, 3. Знайти обсяг вибірки, скласти дискретний статистичний розподіл частот та відносних частот, побудувати полігон частот та відносних частот.

1.3. У колл-центрі інтернет-магазину досліджувалась кількість неправильних з'єднань за хвилину. Спостереження протягом 30 хвилин дали такі результати: 2, 3, 0, 5, 2, 0, 0, 1, 3, 2, 2, 5, 3, 0, 2, 3, 2, 3, 3, 4, 0, 5, 3, 4, 3, 2, 1, 2, 2, 4. Потрібно: 1) скласти статистичний розподіл частот і відносних частот; 2) побудувати полігон частот та відносних частот; 3) знайти емпіричну функцію розподілу та побудувати її графік.

1.4. Обстеження величини рівня прибутковості (рентабельності, у %) підприємства протягом року дало результати 99, 101, 103, 97, 99, 100, 80, 82, 80, 102, 101, 103. Потрібно: 1) скласти статистичний розподіл частот і відносних частот; 2) побудувати полігон частот та відносних частот; 3) знайти емпіричну функцію розподілу та побудувати її графік.

1.5. Після установки нового обладнання у кондитерському цеху було вирішено протягом десяти днів перевіряти торти «Маківчанка» на наявність браку. У результаті було отримано вибірку: 1; 2; 0; 0; 2; 3; 0; 1; 0; 0. Потрібно: 1) скласти статистичний розподіл частот і відносних частот; 2) побудувати полігон частот та відносних частот; 3) знайти емпіричну функцію розподілу та побудувати її графік.

1.6. Результати дослідження річного обсягу споживання олії соняшникової (у кг на одну особу) за регіонами становлять: 11; 12,4; 11; 13,6; 14; 10; 13; 13,6; 12; 12,4; 13,8; 12,9; 13; 14; 12; 10; 11,2; 12,5; 12; 14. Скласти інтервальний статистичний розподіл частот та відносних частот з кроком $h = 2$ і побудувати гістограму частот.

1.7. На підприємстві було проведене підвищення кваліфікації працівників. Після чого було проаналізовано рівень зростання (у %) 15 показників роботи підприємства: 5; 7; 0; 2; 3; 8; 0; 1; 5; 0; 3; 2; 1; 0; 3. Потрібно скласти інтервальний статистичний розподіл частот та відносних частот (кількість інтервалів 4) і побудувати гістограму відносних частот.

1.8. Із генеральної сукупності було отримано вибірку обсягом $n = 30$ зі значеннями: 3; 5; 8; 12; 1; 7; 4; 3; 7; 11; 15; 8; 9; 5; 10; 3; 13; 15; 1; 2; 4; 3; 10; 7;

4; 3; 10; 9; 8; 5. Потрібно: 1) скласти інтервальний статистичний розподіл частот та відносних частот (кількість інтервалів визначити за формулою $k = \left[\sqrt{n}\right]$); 2) побудувати гістограму частот та відносних частот; 3) знайти емпіричну функцію розподілу та побудувати її графік.

1.9. Результати дослідження річного обсягу споживання хліба (у кг на одну особу) за регіонами становлять: 101; 102,5; 101; 103,5; 104; 100; 103; 107,5; 110; 102,5; 108; 105; 103; 104; 102; 107; 101,5; 105; 102; 104; 105,5; 106; 105,5; 107; 108. Потрібно: 1) скласти інтервальний статистичний розподіл частот та відносних частот (кількість інтервалів визначити за формулою Стерджеса); 2) побудувати гістограму частот та відносних частот; 3) знайти емпіричну функцію розподілу та побудувати її графік.

1.10. У відділі перевірки якості сировини олійно-жирового підприємства працює 10 співробітників. Наведено дані про загальний стаж роботи кожного з них: 2; 10; 4; 7; 3; 2; 3; 5; 7; 4. Потрібно: 1) скласти інтервальний статистичний розподіл частот та відносних частот (кількість інтервалів – 4); 2) побудувати гістограму частот та відносних частот; 3) знайти емпіричну функцію розподілу та побудувати її графік.

1.11. Задано статистичний розподіл частот вибірки:

а)

x_i	3	4	5	9
n_i	5	10	7	8

б)

x_i	-2	0	2	4
n_i	2	15	5	8

Знайти розподіл відносних частот, побудувати полігон частот та відносних частот.

1.12. Задано статистичний розподіл частот вибірки:

а)

x_i	1	7	9	12
n_i	1	5	7	2

б)

x_i	5	8	10	14
n_i	7	3	8	2

Знайти емпіричну функцію розподілу та побудувати її графік.

1.13. Задано статистичний розподіл частот вибірки:

а)

$[\alpha_{i-1};\alpha_i)$	$[0;2)$	$[2;4)$	$[4;6)$	$[6;8)$
n_i	5	6	9	5

б)

$[\alpha_{i-1};\alpha_i)$	$[2;7)$	$[7;12)$	$[12;17)$	$[17;22)$
n_i	1	7	10	2

Знайти розподіл відносних частот, побудувати гістограму частот та відносних частот.

1.14. Задано статистичний розподіл частот вибірки:

а)

$[\alpha_{i-1};\alpha_i)$	$[-2;0)$	$[2;4)$	$[4;6)$
n_i	2	3	5

б)

$[\alpha_{i-1};\alpha_i)$	$[0;7)$	$[7;14)$	$[14;21)$	$[21;28)$
n_i	3	6	10	1

Знайти: 1) розподіл відносних частот, 2) емпіричну функцію розподілу та побудувати її графік.

1.15. Задано статистичний розподіл частот вибірки:

а)

$[\alpha_{i-1};\alpha_i)$	$[-1;0)$	$[0;1)$	$[1;2)$	$[2;3)$
n_i	2	8	5	5

б)

$[\alpha_{i-1};\alpha_i)$	$[0;4)$	$[4;8)$	$[8;12)$	$[12;16)$
n_i	1	9	8	2

Потрібно побудувати: 1) гістограму відносних частот; 2) полігон частот; 3) емпіричну функцію розподілу.

Лекція 2. Основні числові характеристики статистичного розподілу вибірки

2.1. Середні вибіркові характеристики

2.2. Характеристики розсіювання

2.1. Середні вибіркові характеристики

Середні величини характеризують значення ознаки, навколо якого концентруються варіанти статистичного розподілу (центральна тенденція розподілу). Існують різні форми середніх: арифметична, гармонічна, квадратична, геометрична тощо. Найбільш поширеною з них є середня арифметична.

Розглянемо дискретний статистичний розподіл вибірки:

x_i	x_1	x_2	$\cdots$	x_k
n_i	n_1	n_2	$\cdots$	n_k

<u>*Означення.*</u> *Вибірковою середньою* (середнє арифметичне значення) такого розподілу називається величина, яка обчислюється за формулою

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i n_i. \qquad (2.1)$$

Очевидно, що $\overline{x}_в = \sum_{i=1}^{k} x_i w_i,$, де w_i – відносні частоти.

За вибіркове середнє неперервного статистичного розподілу приймається вибіркове середнє відповідного дискретного розподілу.

Основні властивості вибіркової середньої.

1^0. Якщо всі варіанти помножити на одне і теж число λ: $u_i = \lambda x_i$, то вибіркова середня збільшиться у стільки ж разів:

$$\overline{u}_в = \frac{1}{n}\sum_{i=1}^{k} u_i n_i = \frac{1}{n}\sum_{i=1}^{k}(\lambda x_i)n_i = \lambda \overline{x}_в.$$

2^0. Якщо всі варіанти змінити на одне і те ж число C: $u_i = x_i \pm C$, то вибіркова середня зміниться на таке ж число:

$$\overline{u}_в = \frac{1}{n}\sum_{i=1}^{k} u_i n_i = \frac{1}{n}\sum_{i=1}^{k}(x_i \pm C)n_i = \overline{x}_в \pm C.$$

3^0. Сума всіх відхилень значень ознаки від вибіркової середньої дорівнює нулю:

$$\sum_{i=1}^{k}(x_i - \overline{x}_в)n_i = 0.$$

Обчислення вибіркової середньої безпосередньо за формулою (2.1) часто приводить до громіздких арифметичних розрахунків. Цього можна уникнути, перейшовши від варіант x_i до умовних варіант u_i:

$$u_i = \frac{x_i - C}{\lambda}, \tag{2.2}$$

де C, λ – відповідно підібрані сталі. Використовуючи властивості $1^0 - 2^0$:

$$\overline{x}_в = \lambda \overline{u}_в + C. \tag{2.3}$$

Іноді при дослідженні сукупності її доводиться розбивати на окремі групи або, навпаки, об'єднувати декілька груп в одну сукупність.

Означення. Груповою середньою називається середнє арифметичне значень ознаки, що належать заданій групі.

Нехай сукупність обсягу n складається із m груп обсягами N_j $(j = \overline{1,m})$ кожна: $n = \sum_{j=1}^{m} N_j$. Якщо $\overline{x}_j$ – групові середні, то загальна вибіркова середня $\overline{x}_в$ обчислюється за формулою

$$\overline{x}_в = \frac{1}{n}\sum_{j=1}^{m} \overline{x}_j N_j. \tag{2.4}$$

Для знаходження вибіркової середньої:

використовується статистична функція **СРЗНАЧ(*число1*; [*число2*]; …)**, де *число 1* – обов’язковий аргумент, посилання на клітинку або діапазон, для якого потрібно визначити середнє значення, число 2 – необов’язковий аргумент, додаткові посилання на клітинки або діапазони, для яких потрібно визначити середнє значення. Аргументів може бути не більше 255. Якщо задано дискретний статистичний розподіл, то вибіркова середня розраховується за формулою, з використанням математичної функції **СУММПРОИЗВ(*массив1*; *массив2*;**…**)**, описаної у ч. 1, с. 63;

використовується функція **mean(*x*)**, де x – вектор або матриця з вибіркою випадкових величин. Також вибіркову середню можна розрахувати безпосередньо за формулою.

Приклад 2.1.

Виміряно зріст (у см) 25, навмання відібраних, студентів:

160; 170; 159; 161; 160;
172; 180; 160; 157; 163;
170; 175; 172; 163; 170;
172; 165; 159; 161; 163;
160; 170; 159; 165; 160.

Обчислити середній зріст студентів.

Розв’язання.

Для обчислення середнього зросту за наведеною в умові вибіркою потрібно скласти дискретний статистичний розподіл величини $X = x_i$ – зріст студентів та для полегшення обчислення ввести умовні варіанти $X - 165$:

x_i	157	159	160	161	163	165	170	172	175	180
$x_i - 165$	-8	-6	-5	-4	-2	0	5	7	10	15
n_i	1	3	5	2	3	2	4	3	1	1

За формулою (2.4) та властивістю 2^0:

$$\bar{x}_в = \frac{1}{25}\sum_{i=1}^{10}(x_i - 165)n_i + 165 = \frac{1}{25}((-8)\cdot 1 + (-6)\cdot 3 + (-5)\cdot 5 + (-4)\cdot 2 + (-2)\cdot 3 +$$

$$+0\cdot 2 + 5\cdot 4 + 7\cdot 3 + 10\cdot 1 + 15\cdot 1) + 165 = 0{,}04 + 165 = 165{,}04.$$

Отже, середній зріст студентів 165,04.

І та ІІ способи

	A	B	C	D	E	F	G	H	I	J	K
1											
2	160	170	159	161	160						
3	172	180	160	157	163						
4	170	175	172	163	170						
5	172	165	159	161	163						
6	160	170	159	165	160						
7				вибіркова середня	**165,04**						
8											
9	зріст	157	159	160	161	163	165	170	172	175	180
10	частота	1	3	5	2	3	2	4	3	1	1
11										вибіркова середня	**165,04**
12											
13											

=СРЗНАЧ(A2:E6)

=СУММПРОИЗВ(B9:K9;B10:K10)/СУММ(B10:K10)

І спосіб

$$A := \begin{pmatrix} 160 & 170 & 159 & 161 & 160 \\ 172 & 180 & 160 & 157 & 163 \\ 170 & 175 & 172 & 163 & 170 \\ 172 & 165 & 159 & 161 & 163 \\ 160 & 170 & 159 & 165 & 160 \end{pmatrix}$$

$$\mathrm{mean}(A) = 165.04$$

II спосіб

$$ORIGIN := 1$$

$$XN := \begin{pmatrix} 157 & 1 \\ 159 & 3 \\ 160 & 5 \\ 161 & 2 \\ 163 & 3 \\ 165 & 2 \\ 170 & 4 \\ 172 & 3 \\ 175 & 1 \\ 180 & 1 \end{pmatrix} \qquad xv := \frac{1}{\sum XN^{\langle 2 \rangle}} \cdot \sum_{i=1}^{rows(XN)} \left(XN_{i,1} \cdot XN_{i,2}\right) = 165.04$$

Відповідь: середній зріст студентів 165,04 см.

Для аналізу статистичного розподілу, крім вибіркової середньої, застосовують структурні або порядкові середні. Найбільш поширеними із них є мода і медіана.

Означення. Модою $\bar{M}_o$ статистичного розподілу називається варіанта, яка має найбільшу частоту появи.

Для дискретного варіаційного ряду мода визначається безпосередньо за означенням. Причому, якщо найбільшу частоту появи мають дві або більше варіант, то такий ряд має дві або більше мод; якщо всі варіанти ряду мають однакову частоту, то у таких випадках мода не визначається.

Для інтервального статистичного розподілу знаходиться інтервал $[\alpha_{i-1};\alpha_i)$, який має найбільшу частоту, – *модальний інтервал*. Значення моди обчислюється шляхом лінійної інтерполяції:

$$\bar{M}_o = \alpha_{i-1} + \frac{n_i - n_{i-1}}{2n_i - n_{i-1} - n_{i+1}}(\alpha_i - \alpha_{i-1}), \tag{2.5}$$

де n_i – частота модального інтервала, n_{i-1}– частота домодального інтервала, n_{i+1} – частота післямодального інтервала.

Для інтервального статистичного розподілу простіше моду знайти графічно за гістограмою частот: визначається абсциса точки перетину діагоналей, які з'єднують вершини прямокутника, побудованого на модальному інтервалі, із найближчими вершинами двох прилеглих прямокутників (рис. 2.1).

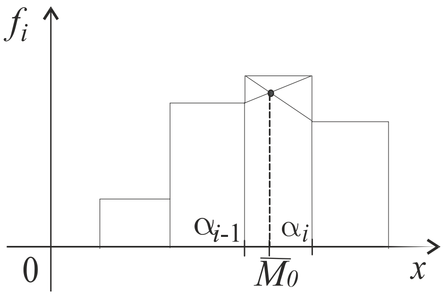


Рис. 2.1

<u>*Означення.*</u> *Медіаною* $\bar{M}_e$ статистичного розподілу називають варіанту, яка припадає на середину розподілу варіаційного ряду.

Для дискретного варіаційного ряду із непарною кількістю елементів медіана дорівнює серединній варіанті, для ряду із парною кількістю елементів – пів сумі двох серединних варіант.

Для інтервального статистичного розподілу визначається *медіанний інтервал* $[\alpha_{i-1};\alpha_i)$ із умови $F_n(\alpha_{i-1})\le 0,5$ і $F_n(\alpha_i)\ge 0,5$. Значення медіани на цьому інтервалі обчислюється шляхом лінійної інтерполяції за формулою:

$$\bar{M}_e=\alpha_{i-1}+\frac{0,5-F_n(\alpha_{i-1})}{w_i}(\alpha_i-\alpha_{i-1}), \tag{2.6}$$

де $w_i=\dfrac{n_i}{n}$.

Медіану наближено можна знайти із графіка емпіричної функції розподілу як аргумент цієї функції, при якому $F_n(x)=0,5$ (рис. 2.2).

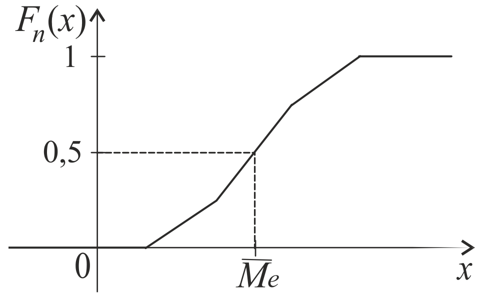


Рис. 2.2

Для знаходження моди та медіани дискретного варіаційного ряду:

використовуються статистичні функції **МОДА(*число1*;[*число2*];...)** та **МЕДІАНА(*число1*;[*число2*];...)**, відповідно, де ***число1,число2,...*** – від 1 до 255 аргументів, для яких обчислюється мода або медіана;

використовуються функції **median(*x*)** та **mode(*x*)**, відповідно, де x – вектор або матриця з вибіркою випадкових величин.

Приклад 2.2.

Наведено інтервальний статистичний розподіл величини X – стаж роботи працівників приватного підприємства:

$[\alpha_{i-1};\alpha_i)$	0-3	3-6	6-9	9-12
n_i	2	7	8	3

Визначити $\overline{M}_o$, $\overline{M}_e$.

Розв'язання.

Обсяг заданої вибірки $n = 20$, найбільшу частоту має інтервал $[6;9)$, отже, він є модальним. Застосувавши формулу (2.5) і врахувавши, що $n_i = 8$, $n_{i-1} = 7$, $n_{i+1} = 3$:

$$\overline{M}_o = 6 + \frac{8-7}{2\cdot 8 - 7 - 3}(9-6) = 6,5.$$

Для обчислення $\bar{M}_e$ потрібно знайти емпіричну функцію розподілу $F_n(x) = W(X < x)$:

$$F_n(x) = \begin{cases} 0 & \text{при } x < 0; \\ 0,1 & \text{при } x = 3; \\ 0,45 & \text{при } x = 6; \\ 0,85 & \text{при } x = 9; \\ 1 & \text{при } x \geq 12. \end{cases}$$

Умова $F_n(\alpha_{i-1}) \leq 0,5$ і $F_n(\alpha_i) \geq 0,5$ виконується для інтервалу $[6;9)$, тому він є медіанним. За формулою (2.6), врахувавши, що $F_n(6) = 0,45$, $w_i = \frac{n_i}{n} = \frac{8}{20} = 0,4$:

$$\bar{M}_e = 6 + \frac{0,5 - 0,45}{0,4}(9 - 6) = 6,375.$$

<u>Відповідь:</u> $\bar{M}_o = 6,5$, $\bar{M}_e = 6,375$.

2.2. Характеристики розсіювання

Середні величини, розглянуті у п. 2.1. ($\bar{x}_в$, $\bar{M}_o$, $\bar{M}_e$), не відображають варіації (розсіювання) значень ознаки. Найпростішою оцінкою розсіювання варіант статистичного розподілу вибірки є *варіаційний розмах*

$$R = x_{\max} - x_{\min}, \tag{2.7}$$

де $x_{\max}$, $x_{\min}$ – найбільше і найменше значення ознаки.

Цей показник залежить від двох крайніх значень ознаки і не враховує проміжних варіант і частот.

З практичної точки зору найбільш важливою є характеристика розсіювання значень варіаційного ряду навколо вибіркової середньої. Найчастіше це розсіювання оцінюється дисперсією і середнім квадратичним відхиленням.

<u>Означення.</u> Дисперсією вибірки $D_в$ називається середнє арифметичне квадратів відхилень варіант відносно вибіркової середньої $\bar{x}_в$:

$$D_в = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_в)^2 n_i \text{ або } D_в = \sum_{i=1}^{k}(x_i - \overline{x}_в)^2 w_i. \quad (2.8)$$

Основні властивості дисперсії вибірки.

1^0. Дисперсія сталої дорівнює нулю.

2^0. Якщо всі варіанти вибірки помножити на одне і те ж число λ, то дисперсія вибірки зміниться у λ^2 разів:

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

3^0. Якщо всі варіанти змінити на одне і те ж число C, то дисперсія вибірки не зміниться:

$$\frac{1}{n}\sum_{i=1}^{k}[(x_i \pm C) - (\overline{x}_в \pm C)]^2 n_i = D_в.$$

4^0. Дисперсія вибірки дорівнює різниці середньої арифметичної квадратів варіант і квадрату вибіркової середньої:

$$D_в = \frac{1}{n}\sum_{i=1}^{k}x_i^2 n_i - (\overline{x}_в)^2 \text{ або } D_в = \left(\overline{x^2}\right)_в - (\overline{x}_в)^2. \quad (2.9)$$

Для спрощення обчислення дисперсій вибірки можна скористатись умовними варіантами (2.2):

$$D_{вх} = \frac{\lambda^2}{n}\sum_{i=1}^{k}u_i^2 n_i - (\overline{x}_в - C)^2 \text{ або } D_{вх} = \lambda^2 D_{ви}, \quad (2.10)$$

де $D_{вх}$, $D_{ви}$ – дисперсії вибірок з варіантами x_i, u_i відповідно.

Нехай усю вибірку обсягу n розбито на m груп. Позначимо через N_j – обсяг j-тої групи, $\overline{x}_j$ – групову середню, $D_{jгр}$ – групову дисперсію ($j = \overline{1, m}$), $\overline{x}_в$ – загальну вибіркову середню, $D_в$ – загальну дисперсію вибірки.

<u>Означення.</u> *Міжгруповою дисперсією* називається величина

$$D_{мгр} = \frac{1}{n}\sum_{j=1}^{m}(\overline{x}_j - \overline{x}_в)^2 N_j. \quad (2.11)$$

Ця величина характеризує дисперсію групових середніх. Загальна дисперсія вибірки обчислюється за формулою

$$D_в = \frac{1}{n}\sum_{j=1}^{m} D_j N_j + D_{мгр}. \qquad (2.12)$$

За оцінку розсіювання бажано мати величину, яка вимірюється в тих же одиницях, що і варіанти ознаки. Такими характеристиками є:

1) *середнє квадратичне відхилення* (або *стандартне відхилення*) вибірки $\sigma_в = \sqrt{D_в}$.

2) коефіцієнт варіації $V = \frac{\sigma_в}{\overline{x}_в} \cdot 100\%$.

Для знаходження характеристик розсіювання:

використовуються статистичні функції **МАКС(*число1*;[*число2*];...)** та **МІН(*число1*;[*число2*];...)** – повертають найбільше та найменше значення, відповідно, із списку аргументів, де ***число1,число2,...*** – від 1 до 255 аргументів, серед яких потрібно знайти найменше або найбільше; **СРЗНАЧ**(***число1***;[число**2**];**...**) та **СРЗНАЧА**(***значение1***; [***значение2***];**...**) – повертають середнє арифметичне вказаних аргументів (відмінність полягає у тому, що перша функція ігнорує логічні і текстові значення, друга – переводить їх у цифровий формат (0 або 1)), **ДИСП**(***число1***;[число**2**];**...**) та **ДИСПА**(***значение1***; [***значение2***];**...**) – оцінюють дисперсію за вибіркою (відмінність аналогічна попереднім функціям), **СТАНДОТКЛОН**(***число1***; [число**2**];**...**) та **СТАНДОТКЛОНА**(***значение1***; [***значение2***];**...**) – оцінюють середнє квадратичне відхилення за вибіркою (відмінність аналогічна попереднім функціям); де ***число1,число2,...*** та ***значение1,значение2,...*** – від 1 до 255 числових аргументів, які відповідають вибірці. Характеристики розсіювання можна розрахувати безпосередньо за формулами (це актуально, якщо вибірка задана у вигляді статистичного розподілу);

використовуються функції **max(*x*)** та **min(*x*)** – максимальне та мінімальне значення вибірки; **var(*x*)**, **stdev(*x*)** – вибіркові дисперсія та середнє квадратичне відхилення, відповідно, де x – вектор або матриця з вибіркою випадкових величин. Характеристики розсіювання можна розрахувати безпосередньо за формулами (це актуально, якщо вибірка задана у вигляді статистичного розподілу).

<u>Приклад 2.3.</u>

Задано інтервальний статистичний розподіл випадкової величини X – кількість балів, одержаних абітурієнтами на вступному іспиті з математики:

x_i	0-30	30-50	50-70	70-100
n_i	10	60	25	5

Обчислити $\overline{x}_в,\ D_в,\ \sigma_в,\ V$.

<u>Розв’язання.</u>

Для обчислення потрібних числових характеристик потрібно перейти від інтервального статистичного розподілу до дискретного, варіантами якого будуть $x_i = \dfrac{\alpha_{i-1} + \alpha_i}{2}$:

x_i	15	40	60	85
n_i	10	60	25	5

Числові характеристики $\overline{x}_в,\ D_в,\ \sigma_в,\ V$ обчислюються за формулами (2.4), (2.8) та характеристиками розсіювання 1), 2) відповідно:

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{1}{100}(15\cdot 10 + 40\cdot 60 + 60\cdot 25 + 85\cdot 5) = 44,75;$$

$$D_в = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_в)^2 n_i = \frac{1}{100}((15-44,75)^2\cdot 10 + (40-44,75)^2\cdot 60 +$$

$$+(60-44,75)^2\cdot 25 + (85-44,75)^2\cdot 5) \approx 241,19;$$

$$\sigma_в = \sqrt{D_в} = \sqrt{241,19} \approx 15,53;$$

$$V = \frac{\sigma_в}{\overline{x}_в}\cdot 100\% = \frac{15,53}{44,75}\cdot 100\% \approx 0,35\cdot 100\% = 35\%.$$

	A	B	C	D	E	F	G	H	I	J	K	L
1	α_{i-1}	α_i	$x_i = \frac{\alpha_{i-1}+\alpha_i}{2}$	n_i								
2	0	30	15	10	$\overline{x}_в$	44,75						
3	30	50	40	60	$D_в$	241,19						
4	50	70	60	25	$\sigma_в$	15,53						
5	70	100	85	5	V	34,704	%					
6			обсяг:	100								
7												
8												

=СУММПРОИЗВ(C2:C5;D2:D5)/D6

=СУММПРОИЗВ(СТЕПЕНЬ(C2:C5-F2;2);D2:D5)/D6

=КОРЕНЬ(F3)

=(F4/F2)*100

=СУММ(D2:D5)

ORIGIN := 1

$$IN := \begin{pmatrix} 0 & 30 & 10 \\ 30 & 50 & 60 \\ 50 & 70 & 25 \\ 70 & 100 & 5 \end{pmatrix} \qquad n := \sum IN^{\langle 3\rangle} = 100$$

$$X := \left| \begin{array}{l} \text{for } i \in 1..\text{rows}(IN) \\ \quad x_{i,1} \leftarrow \dfrac{IN_{i,1} + IN_{i,2}}{2} \\ x \end{array} \right. \qquad X = \begin{pmatrix} 15 \\ 40 \\ 60 \\ 85 \end{pmatrix}$$

$$xv := \frac{1}{n}\cdot \sum_{i=1}^{\text{rows}(X)} \left(X_{i,1}\cdot IN_{i,3}\right) = 44.75$$

$$Dv := \frac{1}{n} \cdot \sum_{i=1}^{rows(X)} \left[\left(X_{i,1} - xv \right)^2 \cdot IN_{i,3} \right] = 241.188$$

$$\sigma v := \sqrt{Dv} = 15.53$$

$$V := \frac{\sigma v}{xv} \cdot 100 = 34.704$$

Відповідь: $\bar{x}_в = 44,75;\ D_в \approx 241,19;\ \sigma_в \approx 15,53;\ V \approx 35\%.$

Приклад 2.4.

За рівнем успішності з предмету «Менеджмент» студенти двох груп розподілені наступним чином:

Рівень успішності	**Група 1**	**Група 2**
4 (високий)	2	1
3 (достатній)	8	10
2 (середній)	5	6
1 (низький)	4	3

Знайти загальну середню, міжгрупову та загальну дисперсії.

Розв’язання.

Групові середні:

$$\bar{x}_1 = \frac{4 \cdot 2 + 3 \cdot 8 + 2 \cdot 5 + 1 \cdot 4}{2 + 8 + 5 + 4} \approx 2,4;\ \ \bar{x}_2 = \frac{4 \cdot 1 + 3 \cdot 10 + 2 \cdot 6 + 1 \cdot 3}{1 + 10 + 6 + 3} \approx 2,5.$$

Загальна вибіркова середня обчислюється за формулою (2.4):

$$\bar{x}_в = \frac{2,4 \cdot 19 + 2,5 \cdot 20}{19 + 20} \approx 2,5.$$

Групові дисперсії:

$$D_1 = \frac{(4 - 2,4)^2 \cdot 2 + (3 - 2,4)^2 \cdot 8 + (2 - 2,4)^2 \cdot 5 + (1 - 2,4)^2 \cdot 4}{19} \approx 0,9;$$

$$D_2 = \frac{(4-2,5)^2 \cdot 1 + (3-2,5)^2 \cdot 10 + (2-2,5)^2 \cdot 6 + (1-2,5)^2 \cdot 3}{19} \approx 0,7.$$

Міжгрупова дисперсія обчислюється за формулою (2.11):

$$D_{мгр} = \frac{(2,4-2,5)^2 \cdot 19 + (2,5-2,5)^2 \cdot 20}{39} \approx 0,05.$$

Загальна дисперсія обчислюється за формулою (2.12):

$$D_в = \frac{0,9 \cdot 19 + 0,7 \cdot 20}{39} + 0,05 \approx 0,85.$$

Відповідь: $\overline{x}_в \approx 2,5;\ D_{мгр} \approx 0,05;\ D_в \approx 0,85.$

Практичне заняття 2

2.1. Молокозавод постачає йогурти у дитячий садочок. Результати замовлень (у шт. по 200 мг) протягом місяця становлять: 150; 120; 110; 145; 140; 135; 122; 126; 145; 138; 135; 135; 147; 140; 140; 140; 145; 145; 130; 130. Знайти середню кількість замовлень протягом місяця.

2.2. Наведено результати опитування мешканців деякого регіону щодо споживання овочів та баштанних (у кг на одну особу) протягом літа: 30; 28; 45; 30; 45; 27; 15; 35; 39; 40; 25; 30; 45; 35; 25; 34; 22; 36; 27; 48. Потрібно побудувати дискретний статистичний розподіл вибірки та знайти її основні числові характеристики ($\overline{x}_в,\ D_в,\ \sigma_в$).

2.3. Задано статистичний розподіл вибірки:

а)

x_i	-3,5	0,5	4,5
n_i	3	10	7

б)

x_i	0,001	0,005	0,007	0,009
n_i	3	7	10	5

Обчислити $\overline{x}_в,\ D_в,\ \sigma_в,\ V$.

2.4. Задано статистичний розподіл вибірки:

а)

x_i	-3	1	5
n_i	5	10	5

б)

x_i	1	4	7	10
n_i	2	8	7	3

Визначити $\overline{M}_o,\ \overline{M}_e$.

2.5. Статистичний розподіл врожайності соняшнику (у т/га) за деяким регіоном подано у вигляді таблиці:

x_i, т/га	1,5	1,9	2	2,3
n_i	5	10	11	4

Потрібно: 1) обчислити $\overline{x}_в,\ D_в,\ \sigma_в,\ V$; 2) визначити $\overline{M}_o,\ \overline{M}_e$.

2.6. Задано статистичний розподіл вибірки:

а)

$\alpha_{i-1}-\alpha_i$	0-3	3-6	6-9
n_i	1	9	5

б)

$\alpha_{i-1}-\alpha_i$	1-6	6-11	11-16	16-21
n_i	4	5	7	4

Обчислити $\overline{x}_в,\ D_в,\ \sigma_в,\ V$.

2.7. Задано статистичний розподіл вибірки:

а)

$[\alpha_{i-1};\alpha_i)$	$[-3;0)$	$[0;3)$	$[3;6)$
n_i	5	10	5

б)

$[\alpha_{i-1};\alpha_i)$	$[-10;0)$	$[0;10)$	$[10;20)$	$[20;30)$
n_i	2	7	11	10

Визначити $\bar{M}_o,\ \bar{M}_e$.

2.8. Вивчались показники інноваційної діяльності 50 підприємств харчової промисловості деякого регіону. Дані про обсяг реалізованої інноваційної продукції подані у вигляді статистичного розподілу:

Інтервал, тис.ум.од.	1-3	3-5	5-7	7-9
n_i	15	23	7	5

Потрібно: 1) обчислити $\bar{x}_в,\ D_в,\ \sigma_в,\ V$; 2) визначити $\bar{M}_o,\ \bar{M}_e$.

2.9. Наведено результати опитування мешканців міста Львова щодо відвідування кав'ярень (у разах) протягом місяця: 5; 4; 3; 25; 20; 5; 15; 35; 7; 30; 25; 3; 20; 30; 27; 17; 19; 21; 9; 11; 9; 12; 11; 21; 3; 0; 0; 5; 2; 22. Потрібно побудувати інтервальний статистичний розподіл вибірки з кроком $h=10$ та знайти $\bar{x}_в,\ D_в,\ \sigma_в,\ V,\ \bar{M}_o,\ \bar{M}_e$.

2.10. Мобільний оператор «All» провів аналіз скарг абонентів (тис. осіб) протягом року: 1,1; 3,1; 2; 5,5; 0,5; 7; 3,5; 0,9; 1,5; 1,3; 2,3; 3,2. Потрібно побудувати інтервальний статистичний розподіл вибірки з кроком $h=2$ та знайти $\bar{x}_в,\ D_в,\ \sigma_в,\ V,\ \bar{M}_o,\ \bar{M}_e$.

2.11. Сукупність складена з двох груп:

1)

x_i	0	1	3
n_i	3	12	3

2)

x_i	-2	0	4
n_i	10	11	4

Знайти групові та загальну вибіркову середні.

2.12. Сукупність складена з двох груп:

1)

$\alpha_{i-1} - \alpha_i$	0-1	1-2	2-3
n_i	2	15	3

2)

$\alpha_{i-1} - \alpha_i$	1-5	5-9	9-13
n_i	8	2	3

Знайти групові та загальну вибіркову середні.

2.13. Вивчались дані про стаж роботи працівників у концерні «Веспром». Результати подано у вигляді таблиці:

Стаж роботи, років	**Під-во «Ковбаска»**	**Під-во «Молочник»**	**Під-во «Соняшник»**
0-2	4	8	10
2-5	25	20	10
5-8	5	8	10
8-10	3	5	2
10-15	0	2	1

Обчислити: 1) середній стаж роботи по кожному із підприємств і у концерні загалом; 2) групові, міжгрупову і загальну дисперсії.

2.14. Вивчались дані про успішність потоку студентів з дисципліни «Інноваційний менеджмент». Результати подано у вигляді таблиці:

Оцінка	**Група 1**	**Група 2**	**Група 3**
1	**2**	**3**	**4**
5	5	3	5
4	11	9	10
1	**2**	**3**	**4**
3	2	8	3
2	1	1	2
1	0	1	0

Обчислити: 1) середній бал по кожній з груп; 2) загальний середній бал за потоком; 3) групові, міжгрупову і загальну дисперсії.

Лекція 3. Статистичні оцінки параметрів генеральної сукупності

3.1. Поняття статистичних оцінок і загальні вимоги до них

Нехай досліджується випадкова величина X генеральної сукупності обсягу N. Виникає задача за результатами однієї вибірки оцінити невідомі параметри генеральної сукупності, такі як, наприклад, математичне сподівання, дисперсію, ймовірність тощо.

Виділяється два типи задач про оцінку параметрів: якщо величину, обчислену за вибіркою, можна взяти за наближене значення характеристики генеральної сукупності (*точкова оцінка*); в якому проміжку з заданою надійністю буде знаходитися істинна характеристика (*інтервальна оцінка*).

Випадковий відбір дозволяє вибірку $x_1, x_2, ..., x_n$ обсягу n розглядати як часткові значення (реалізацію) n незалежних випадкових величин $X_1, X_2, ..., X_n,$ кожна із яких має такий самий закон розподілу ймовірностей, що і ознака X генеральної сукупності, з відповідними числовими характеристиками.

<u>Означення.</u> Статистичною оцінкою невідомого параметра θ генеральної сукупності називається функція $\theta_n^* = \theta^*(X_1, X_2, \ldots, X_n)$ вибіркових значень

$X_1, X_2, ..., X_n$, яка в деякому статистичному розумінні близька до істинного значення θ.

Функція θ_n^* залежить від випадкових величин $X_1, X_2, ..., X_n$ тому вона теж є випадковою величиною, що залежить від закону розподілу ознаки генеральної сукупності X та числа n.

До статистичних оцінок висуваються вимоги *незсуненості*, *ґрунтовності* і *ефективності*, які дозволяють знайти оцінку найбільш близьку до параметра, який оцінюється.

Означення. Оцінка θ_n^* називається *незсуненою*, якщо математичне сподівання цієї оцінки точно дорівнює оцінювальному параметру, тобто

$$M(\theta_n^*) = \theta. \tag{3.1}$$

Інакше оцінка називається *зсуненою*. Вимога незсуненості гарантує відсутність систематичних помилок при оцінюванні.

Означення. Оцінка θ_n^* називається *ґрунтовною*, якщо у разі необмеженого збільшення обсягу вибірки оцінка θ_n^* наближається до оцінювального параметра θ, тобто для довільного $\varepsilon > 0$ виконується гранична рівність

$$\lim_{n\to\infty} P\left\{\left|\theta_n^* - \theta\right| < \varepsilon\right\} = 1. \tag{3.2}$$

Ця властивість виражає закон великих чисел щодо оцінки θ_n^*: при достатньо великому n достовірною є рівність $\theta_n^* \approx \theta$.

Означення. Оцінка θ_n^* називається *ефективною*, якщо при заданому обсязі вибірки вона має мінімальну дисперсію серед усіх інших оцінок.

Приведена властивість свідчить про якість використаної оцінки.

3.2. Методи визначення точкових статистичних оцінок

Серед методів визначення точкових статистичних оцінок найчастіше використовуються методи аналогій і найменших квадратів.

Метод аналогій. Цей метод ґрунтується на тому, що для оцінок параметрів генеральної сукупності вибираються аналогічні параметри вибірки. Для оцінки частки ознаки $p_i = \frac{N_i}{N}$, генеральної середньої $\overline{X}_Г = \frac{1}{N}\sum_i x_i n_i$ і генеральної дисперсії $D_Г = \frac{1}{N}\sum_k (x_k - \overline{X}_Г)^2 n_k$ вибирають відповідні статистики θ_n^*: вибіркову частку $w = \frac{m}{n}$, вибіркову середню $\overline{x}_в = \frac{1}{n}\sum_i x_i n_i$, вибіркову дисперсію $D_в = \frac{1}{n}\sum_k (x_k - \overline{x}_в)^2 n_i$.

Метод найменших квадратів. Згідно з цим методом статистичні оцінки визначаються з умови мінімізації суми квадратів відхилень варіант вибірки від статистичної оцінки θ_n^*.

Нехай, наприклад, необхідно знайти оцінку θ_n^* для генеральної середньої $\overline{X}_Г = M(X)$.

Згідно розглядуваного методу її слід визначати із умови мінімізації виразу $u = \sum_{i=1}^{n}(x_i - \theta_n^*)^2 n_i$. За необхідною умовою екстремуму: $\frac{\partial u}{\partial \theta_n^*} = -2\sum_{i=1}^{n}(x_i - \theta_n^*) n_i = 0$, звідси $\sum_{i=1}^{n} x_i n_i - \sum_{i=1}^{n}\theta_n^* n_i = 0$ $\left(\sum_{i=1}^{n} n_i = n\right)$ або $\theta_n^* = \frac{1}{n}\sum_i x_i n_i = \overline{x}_в$.

Отже, для $\theta = \overline{X}_Г$ точковою статистичною оцінкою буде вибіркова середня $\overline{x}_в$.

3.3. Точкові оцінки математичного сподівання та дисперсії генеральної сукупності

Нехай $\overline{X}_Г, D_Г$ – середня і дисперсія генеральної сукупності відповідно; $\overline{x}_в$, $D_в$ вибіркова середня і дисперсія, знайдені за вибіркою $x_1, x_2, \ldots, x_n$ обсягу *n*. Слід нагадати, що вибірку можна розглядати як сукупність незалежних випадкових величин $X_1, X_2, \ldots, X_n$, які мають такий самий закон розподілу, що і ознака X генеральної сукупності, причому $M(X_i) = M(X)$.

Точковими незсуненими статистичними оцінками для $\bar{X}_Г = M(X)$ і $D_Г$ є $\bar{x}_в$ і $S^2 = \frac{n}{n-1} D_в$.

Дійсно, оскільки $\bar{x}_в = \bar{X}_в = \frac{1}{n}\sum_{i=1}^{n} X_i$, то

$$M(\bar{X}_в) = M\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\sum_{i=1}^{n} M(X_i) = \frac{1}{n}\cdot n \cdot M(X) = \bar{X}_Г.$$

Отже, $M(\bar{X}_в) = \bar{X}_Г$.

За статистичну оцінку генеральної дисперсії $D_Г = \sigma^2$ потрібно взяти величину

$$D_в = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X}_в)^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X}_в)^2.$$

Оскільки $(\bar{X}_в)^2 = \frac{1}{n^2}\left(\sum_{i=1}^{n} X_i\right)^2 = \frac{1}{n^2}\left(\sum_{i=1}^{n} X_i^2 + \sum_{i \neq j} X_i X_j\right)$,

то $D_в = \frac{n-1}{n^2}\sum_{i=1}^{n} X_i^2 - \frac{1}{n^2}\sum_{i \neq j} X_i X_j$.

Від усіх X_i або X_j слід відняти сталу величину $\mu = \bar{X}_Г$ (при цьому дисперсія не зміниться) і обчислити математичне сподівання обох частин останньої рівності:

$$M(D_в) = \frac{n-1}{n^2}\sum_{i=1}^{n} M(X_i - \mu)^2 - \frac{1}{n^2}\sum_{i \neq j} M\left[(X_i - \mu)(X_j - \mu)\right].$$

Враховуючи, що $M(X_i - \mu)^2 = \sigma^2$, а для незалежних випадкових величин $X_i, X_j \ (i \neq j)$ $M\left[(X_i - \mu)(X_j - \mu)\right] = M(X_i - \mu) M(X_j - \mu) = 0$:

$$M(D_в) = \frac{n-1}{n}\sigma^2 = \frac{n-1}{n} D_Г.$$

Звідси видно, що $D_в$ є зсуненою статистичною оцінкою для $D_Г$.

Для величини

$$S^2 = \frac{n}{n-1} D_в \qquad (3.3)$$

математичне сподівання $M(S^2) = M\left(\frac{n}{n-1} D_в\right) = \frac{n}{n-1} \cdot \frac{n-1}{n} D_Г = D_Г.$

Отже, S^2 – незсунена статистична оцінка для $D_Г$.

Таку оцінку називають *виправленою дисперсією*. Величину

$$S = \sqrt{\frac{n}{n-1} D_в} \qquad (3.4)$$

називають *виправленим середнім квадратичним відхиленням*.

В силу закону великих чисел (теорема Чебишова) статистичні оцінки $\bar{x}_в$ і S^2 будуть ґрунтовними для $\bar{X}_Г$ і $D_Г$.

Отже, отримано оцінки: $\bar{X}_Г \approx \bar{x}_в$, $D_Г \approx S^2$.

Для знаходження точкових незсунених оцінок математичного сподівання та дисперсії генеральної сукупності:

- використовуються функції **ДИСПР(*число1*;**[число**2**]**;...)** та **ДИСПРА(*значение1*;** [***значение2***]**;...)** – оцінюють дисперсію генеральної сукупності (відмінність полягає у тому, що перша функція ігнорує логічні і текстові значення, друга – переводить їх у цифровий формат (0 або 1)). Точкові оцінки математичного сподівання та дисперсії генеральної сукупності можна розрахувати безпосередньо за формулами (це актуально, якщо вибірка задана у вигляді статистичного розподілу);
- використовуються функції **Var(*x*)**, **Stdev(*x*)**, де x – вектор або матриця з вибіркою випадкових величин (потрібно відмітити, що у даному випадку важливим є написання вказаних функцій з великої літери). Точкові оцінки математичного сподівання та дисперсії генеральної

сукупності можна розрахувати безпосередньо за формулами (це актуально, якщо вибірка задана у вигляді статистичного розподілу).

Приклад 3.1.

Знайти точкові незсунені статистичні оцінки генеральних середньої та дисперсії за дискретним статистичним розподілом вибірки:

x_i	1	2	5	8
n_i	3	7	10	5

Розв’язання.

Для обчислення точкових незсунених статистичних оцінок генеральних середньої та дисперсії $\overline{X}_Г,\ D_Г$ потрібно знайти $\overline{x}_в, D_в$ за формулами (2.4), (2.8):

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{1}{25}(1\cdot 3 + 2\cdot 7 + 5\cdot 10 + 8\cdot 5) = 4,28;$$

$$D_в = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_в)^2 n_i = \frac{1}{25}((1-4,28)^2\cdot 3 + (2-4,28)^2\cdot 7 +$$

$$+(5-4,28)^2\cdot 10 + (8-4,28)^2\cdot 5) \approx 5,72.$$

Тоді точковими незсуненими статистичними оцінками для генеральної середньої та дисперсії $\overline{X}_Г,\ D_Г$ є величини

$$\overline{X}_Г \approx \overline{x}_в = 4,28;$$

$$D_Г \approx S^2 = \frac{n}{n-1}\cdot D_в = \frac{25}{24}\cdot 5,72 \approx 5,96.$$

	A	B	C	D
1	x_i	n_i		
2	1	3	$\overline{X}_Г$	4,28
3	2	7		
4	5	10	$D_Г$	5,96
5	8	5		
6	**обсяг:**	**25**		
7				

=СУММПРОИЗВ(A2:A5;B2:B5)/B6

=СУММПРОИЗВ(СТЕПЕНЬ(C2:C5-F2;2);D2:D5)/D6

=СУММ(B2:B5)

$$ORIGIN := 1$$

$$XN := \begin{pmatrix} 1 & 3 \\ 2 & 7 \\ 5 & 10 \\ 8 & 5 \end{pmatrix} \qquad n := \sum XN^{\langle 2 \rangle} = 25$$

$$xv := \frac{1}{n} \cdot \sum_{i=1}^{rows(XN)} \left(XN_{i,1} \cdot XN_{i,2} \right) = 4.28$$

$$Dv := \frac{1}{n} \cdot \sum_{i=1}^{rows(XN)} \left[\left(XN_{i,1} - xv \right)^2 \cdot XN_{i,2} \right] = 5.722$$

$$S2 := \frac{n}{n-1} \cdot Dv = 5.96$$

Відповідь: точкові незсунені статистичні оцінки генеральних середньої та дисперсії: $\overline{X}_Г \approx 4,28$; $D_Г \approx 5,96$.

3.4. Інтервальні оцінки для математичного сподівання нормально розподіленої генеральної сукупності

Точкові статистичні оцінки θ_n^* є випадковими величинами, а тому наближена заміна θ на θ_n^* може привести до суттєвих похибок, особливо, коли обсяг вибірки малий. Тому доцільно точкову оцінку доповнити інтервальною статистичною оцінкою.

Нехай для випадкової величини $|\theta - \theta_n^*|$ виконується нерівність

$$|\theta - \theta_n^*| < \delta, \delta > 0. \tag{3.5}$$

Звичайно, чим менше δ, тим точніша оцінка θ_n^*. Величина δ називається *точністю оцінки*. Оскільки θ_n^* – випадкова величина, то нерівність (3.5) справедлива з певною ймовірністю γ.

Означення. Надійністю оцінки θ_n^* параметра θ називається ймовірність γ, з якою виконується нерівність $|\theta - \theta_n^*| < \delta$, отже, $P(|\theta - \theta_n^*| < \delta) = \gamma$.

Як правило γ задають наперед і беруть близьким до 1 (наприклад, 0,95; 0,99; 0,999).

Означення. Надійним інтервалом називається інтервал $(\theta_n^* - \delta; \theta_n^* + \delta)$, який покриває невідомий параметр θ із заданою ймовірністю γ.

Для знаходженні надійного інтервалу для оцінки математичного сподівання $M(X) = \mu$ потрібно розглянути два випадки, коли відоме середнє квадратичне відхилення σ і коли воно невідоме.

Надійний інтервал для оцінки математичного сподівання $M(X) = \mu$ *нормального розподілу при відомому середньому квадратичному відхиленні* σ.

Нехай ознака X генеральної сукупності розподілена нормально і відоме середнє квадратичне відхилення σ. Для того, щоб за вибірковим середнім $\overline{x}_в$ оцінити невідоме математичне сподівання $M(X) = \mu$, потрібно знайти надійний інтервал, який містить параметр μ із заданою ймовірністю γ.

Оскільки, $\overline{x}_в$ – нормально розподілена випадкова величина, то її числові характеристики:

$$M(\overline{x}_в) = M\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} M(X_i) = \frac{1}{n}\cdot n \cdot \mu = \mu;$$

$$\sigma(\overline{x}_в) = \sqrt{D(\overline{x}_в)} = \sqrt{D\left(\frac{X_1 + X_2 + \ldots + X_n}{n}\right)} = \sqrt{\frac{1}{n^2}\cdot n \cdot D(X)} = \frac{\sigma}{\sqrt{n}}.$$

Для нормально розподіленої випадкової величини X з параметрами μ і σ справедлива формула: $P(|X - \mu| < \delta) = 2\Phi\left(\frac{\delta}{\sigma}\right)$, звідси

$$P(|\overline{x}_в - \mu| < \delta) = 2\Phi\left(\frac{\delta}{\sigma(\overline{x}_в)}\right). \tag{3.6}$$

Оскільки $\sigma(\overline{x}_в) = \sqrt{D(\overline{x}_в)} = \frac{\sigma}{\sqrt{n}}$, то із рівності (3.6):

$$P(|\overline{x}_в - \mu| < \delta) = 2\Phi\left(\frac{\delta\sqrt{n}}{\sigma}\right) = \gamma.$$

Якщо позначити $t = \frac{\delta\sqrt{n}}{\sigma}$, то $\delta = \frac{t\sigma}{\sqrt{n}}$ і попередня формула матиме вигляд:

$$P\left(\overline{x}_в - \frac{t\sigma}{\sqrt{n}} < \mu < \overline{x}_в + \frac{t\sigma}{\sqrt{n}}\right) = 2\Phi(t) = \gamma.$$

Отже, *надійний інтервал для математичного сподівання* μ *при відомому середньому квадратичному відхиленні* σ та заданій надійності γ матиме вигляд:

$$\overline{x}_в - \frac{t\sigma}{\sqrt{n}} < \mu < \overline{x}_в + \frac{t\sigma}{\sqrt{n}}, \tag{3.7}$$

де t знаходиться за таблицею значень функції Лапласа при умові, що $\Phi(t) = \frac{\gamma}{2}$.

При відомих γ і σ мінімальний обсяг вибірки, який забезпечує задану точність δ, визначається за формулою:

$$n_{\min} = \left[\left(\frac{t\sigma}{\delta}\right)^2\right], \tag{3.8}$$

де [...] – ціла частина числа $\left(\frac{t\sigma}{\delta}\right)^2$.

Зауваження. Якщо обсяг вибірки n великий, то інтервальну оцінку (3.7) можна використовувати і за відсутності нормального розподілу величини X.

Надійний інтервал для оцінки математичного сподівання $M(X)=\mu$ *нормального розподілу при невідомому середньому квадратичному відхиленні* σ.

Для знаходженні інтервальної оцінки для μ, якщо σ невідоме, потрібно розглянути випадкову величину t, яка має розподіл Стьюдента з $k=n-1$ ступенями свободи:

$$t=\frac{\overline{x}_в-\mu}{S/\sqrt{n}}, \tag{3.9}$$

де n – обсяг вибірки, $S=\sqrt{\frac{n}{n-1}D_в}$.

Якщо задано надійність γ, то

$$P\left(\frac{|\overline{x}_в-\mu|}{S/\sqrt{n}}<t_\gamma\right)=\gamma \text{ або } P\left(\overline{x}_в-\frac{S\cdot t_\gamma}{\sqrt{n}}<\mu<\overline{x}_в+\frac{S\cdot t_\gamma}{\sqrt{n}}\right)=\gamma.$$

Отже, *надійний інтервал для математичного сподівання* μ *при невідомому середньому квадратичному відхиленні* σ та заданій надійності γ матиме вигляд:

$$\overline{x}_в-\frac{S\cdot t_\gamma}{\sqrt{n}}<\mu<\overline{x}_в+\frac{S\cdot t_\gamma}{\sqrt{n}}, \tag{3.10}$$

де t_γ знаходиться за таблицею значень розподілу Стьюдента (додаток 6) при заданих значеннях γ і $k=n-1$: $t_\gamma=t(\gamma,k)$.

Зауваження. При $k=n-1>30$ випадкова величина t (3.9) має розподіл, близький до нормального, тому у такому випадку t_γ знаходиться за таблицею значень функції Лапласа (додаток 3) із умови $\Phi(t_\gamma)=\frac{\gamma}{2}$.

Для знаходження значень розподілу Стьюдента та функції Лапласа:

використовуються статистичні функції **СТЬЮДРАСПОБР(*вероятность*;*степени_свободы*)**, яка повертає значення розподілу Стьюдента, та **НОРМОБР(*вероятность*;*среднее*;*стандартное_откл*)**, яка повертає аргумент функції Лапласа, де *вероятность* – відповідне значення ймовірності, *степени_свободы* – число ступенів свободи, яке характеризує розподіл Стьюдента, *среднее* – середнє арифметичне значення розподілу, *стандартное_откл* – стандартне відхилення розподілу, причому *среднее*=0, *стандартное_откл*=1; також для знаходження значень функції Лапласа можна використати статистичну функцію **НОРМСТОБР(*вероятность*)**, де *вероятность* – відповідне значення ймовірності, причому замість *вероятность* слід підставляти значення збільшене на 0,5;

використовуються функції **qt(*p*,*d*)** і **qnorm(*p*, μ, σ)** відповідно, де p – ймовірність, $d=n-1$, n – обсяг вибірки, μ – середнє арифметичне значення розподілу, σ – стандартне відхилення розподілу, причому $\mu=0$, $\sigma=1$.

<u>Приклад 3.2.</u>

Для обчислення середньої врожайності пшениці x_i поле було поділено на 30 однакових ділянок. Урожайність на кожній ділянці подана у вигляді дискретного статистичного розподілу:

x_i, ц/га	5,5	6	6,5	7	7,5	8
n_i	2	5	7	8	6	2

З надійністю $\gamma = 0{,}98$ визначити надійний інтервал для дійсного середнього значення урожайності пшениці.

Розв'язання.

Оскільки σ невідоме, то для визначення надійного інтервалу для $\overline{X}_\Gamma = \mu$ потрібно використати формулу (3.10):

$$\overline{x}_в - \frac{S \cdot t_\gamma}{\sqrt{n}} < \mu < \overline{x}_в + \frac{S \cdot t_\gamma}{\sqrt{n}}.$$

Для цього потрібно знайти $\overline{x}_в, S$:

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{1}{30}(5{,}5 \cdot 2 + 6 \cdot 5 + 6{,}5 \cdot 7 + 7 \cdot 8 + 7{,}5 \cdot 6 + 8 \cdot 2) \approx 6{,}8;$$

$$D_в = \frac{1}{n}\sum_{i=1}^{k} (x_i - \overline{x}_в)^2 n_i = \frac{1}{30}((5{,}5 - 6{,}8)^2 \cdot 2 + (6 - 6{,}8)^2 \cdot 5 + (6{,}5 - 6{,}8)^2 \cdot 7 +$$

$$+(7 - 6{,}8)^2 \cdot 8 + (7{,}5 - 6{,}8)^2 \cdot 6 + (8 - 6{,}8)^2 \cdot 2) \approx 0{,}4;$$

$$S = \sqrt{\frac{n}{n-1} \cdot D_в} = \sqrt{\frac{30}{29} \cdot 0{,}4} \approx 0{,}7.$$

Для знаходження t_γ потрібно скористатись таблицею значень розподілу Стьюдента при заданих значеннях $\gamma = 0{,}98$ і $k = n - 1 = 30 - 1 = 29$: $t_\gamma = t(0{,}98; 29) = 2{,}462.$

Отже, надійний інтервал для $\overline{X}_\Gamma = \mu$:

$$6{,}8 - \frac{0{,}7 \cdot 2{,}462}{\sqrt{30}} < \mu < 6{,}8 + \frac{0{,}7 \cdot 2{,}462}{\sqrt{30}} \text{ або } 6{,}48 < \mu < 7{,}09.$$

	A	B	C	D	E
1	x_i	n_i			=СУММПРОИЗВ(A2:A7;B2:B7)/B8
2	5,5	2	$\bar{x}_в$	6,78333	
3	6	5			=СУММПРОИЗВ(СТЕПЕНЬ(A2:A7-D2;2);B2:B7)/B8
4	6,5	7	$D_в$	0,44472	
5	7	8			=КОРЕНЬ(СУММПРОИЗВ(СТЕПЕНЬ(A2:A7-D2;2);B2:B7)/(B8-1))
6	7,5	6	S	0,67828	
7	8	2	=СУММ(B2:B7)		
8	**обсяг:**	**30**			
9	γ	0,98	=СТЬЮДРАСПОБР(1-B9;B8-1)		
10	t_γ	2,46202			
11			=D2-(D6*B10/КОРЕНЬ(B8))		
12	μ_1	6,47845			
13			=D2+(D6*B10/КОРЕНЬ(B8))		
14	μ_2	7,08822			

$$ORIGIN := 1$$

$$XN := \begin{pmatrix} 5.5 & 2 \\ 6 & 5 \\ 6.5 & 7 \\ 7 & 8 \\ 7.5 & 6 \\ 8 & 2 \end{pmatrix} \qquad n := \sum XN^{\langle 2 \rangle} = 30$$

$$xv := \frac{1}{n} \cdot \sum_{i=1}^{rows(XN)} \left(XN_{i,1} \cdot XN_{i,2}\right) = 6.783$$

$$Dv := \frac{1}{n} \cdot \sum_{i=1}^{rows(XN)} \left[\left(XN_{i,1} - xv\right)^2 \cdot XN_{i,2}\right] = 0.445$$

$$S := \sqrt{\frac{n}{n-1} \cdot Dv} = 0.678$$

$$\gamma := 0.98 \qquad \alpha := 1 - \gamma \qquad t := qt\left(1 - \frac{\alpha}{2}, n - 1\right) = 2.462$$

$$\mu_1 := xv - \frac{S \cdot t}{\sqrt{n}} = 6.478 \qquad \mu_2 := xv + \frac{S \cdot t}{\sqrt{n}} = 7.088$$

Відповідь: надійний інтервал для дійсного середнього значення урожайності пшениці: $6,48 < \mu < 7,09$.

Приклад 3.3.

Після виміру врожайності жита на 36 ділянках було визначено вибіркову середню $\bar{x}_в = 3,7$. Із надійністю $\gamma = 0,95$ побудувати надійний інтервал для дійсного середнього значення урожайності жита $\bar{X}_Г$, якщо середнє квадратичне відхилення $\sigma = 0,5$.

Розв'язання.

Оскільки σ відоме, то для визначення надійного інтервалу для $\bar{X}_Г = \mu$ потрібно використати формулу (3.7):

$$\bar{x}_в - \frac{t\sigma}{\sqrt{n}} < \mu < \bar{x}_в + \frac{t\sigma}{\sqrt{n}}.$$

Величина t знаходиться за таблицею значень функції Лапласа при умові, що $\Phi(t) = \frac{\gamma}{2} = \frac{0,95}{2} = 0,475$, тоді $t = 1,96$.

Отже, надійний інтервал для $\bar{X}_Г = \mu$:

$$3,7 - \frac{1,96 \cdot 0,5}{\sqrt{36}} < \mu < 3,7 + \frac{1,96 \cdot 0,5}{\sqrt{36}} \text{ або } 3,54 < \mu < 3,86.$$

	A	B	C	D	E
1	$\bar{x}_в$	3,7			
2	σ	0,5			
3	γ	0,95			
4	n	36			
5	t	1,959964			
6					
7	μ_1	3,53667			
8					
9	μ_2	3,86333			

=НОРМОБР(B3/2+0,5;0;1)

=B1-(B2*B5/КОРЕНЬ(B4))

=B1+(B2*B5/КОРЕНЬ(B4))

$xv := 3.7 \quad n := 36 \quad \sigma := 0.5 \quad \gamma := 0.95$

$\alpha := 1 - \gamma$

$t := \mathrm{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) = 1.96$

$\mu_1 := xv - \frac{\sigma \cdot t}{\sqrt{n}} = 3.537 \qquad \mu_2 := xv + \frac{\sigma \cdot t}{\sqrt{n}} = 3.863$

Відповідь: надійний інтервал для дійсного середнього значення урожайності жита: $3{,}54 < \mu < 3{,}86$.

Приклад 3.4.

Якого значення має набути надійність оцінки γ, щоб за обсягу вибірки $n = 196$ і $\sigma = 5$ її похибка не перевищувала 0,1.

Розв'язання.

Оскільки похибка вибірки $\delta = \frac{t\sigma}{\sqrt{n}}$, то звідси $t = \frac{\delta\sqrt{n}}{\sigma} = \frac{0{,}1 \cdot \sqrt{196}}{5} = 0{,}28$.

Отже, $\gamma = P\left(\left|\overline{x}_в - \mu\right| < \delta\right) = 2\Phi(t) = 2\Phi(0{,}28) = 0{,}2206$, це свідчить про те, що надійність дуже мала.

$n := 196 \quad \sigma := 5 \quad \delta := 0.1$

$t := 0$

Given

$\delta = \frac{t \cdot \sigma}{\sqrt{n}}$

$\mathrm{Find}(t) = 0.28$

$\gamma := 2 \cdot (\mathrm{pnorm}(0.28, 0, 1) - 0.5) = 0.221$

Примітка. У розрахунках було використано блок **Given-Find** описаний у Ч. I, с. 44 – 45.

Відповідь: надійність $\gamma = 0,2206$.

Приклад 3.5.

Визначити мінімальний обсяг вибірки $n_{\min}$, за якого похибка $\delta = 0,02$ гарантувалася б з ймовірністю $0,99$, якщо $\sigma = 4$.

Розв'язання.

Для знаходження мінімального обсягу вибірки $n_{\min}$ потрібно використати формулу (3.8):

$$n_{\min} = \left[\left(\frac{t\sigma}{\delta} \right)^2 \right].$$

Величина t знаходиться за таблицею значень функції Лапласа при умові, що $\Phi(t) = \frac{\gamma}{2} = \frac{0,99}{2} = 0,495$, тоді $t = 2,576$.

Тоді $n_{\min} = \left[\left(\frac{2,576 \cdot 4}{0,02} \right)^2 \right] = 265431$.

$$\sigma := 4 \qquad \delta := 0.02 \qquad \gamma := 0.99$$

$$\alpha := 1 - \gamma$$

$$t := \mathrm{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) = 2.576$$

$$n := \mathrm{trunc}\left[\left(\frac{t \cdot \sigma}{\delta}\right)^2\right] = 2.654 \times 10^5$$

Примітка. У розрахунках було використано функцію **trunc(*x*)**, яка повертає цілу частину дійсного числа *x*.

Відповідь: мінімальний обсяг вибірки $n_{\min} = 265431$.

3.5. Інтервальна оцінка дисперсії (середнього квадратичного відхилення)

Нехай ознака X генеральної сукупності має нормальний розподіл: $X \sim N(\mu, \sigma)$. Для знаходженні надійного інтервалу для оцінки $D_Г$ і σ потрібно розглянути два випадки, коли відоме математичне сподівання μ і коли воно невідоме.

Надійний інтервал для оцінки дисперсії $D_Г$ і середнього квадратичного відхилення σ нормального розподілу при невідомому математичному сподіванні μ.

Для знаходженні інтервальної оцінки для $D_Г$ і σ при невідому μ, потрібно розглянути випадкову величину $\chi^2(k)$, яка має розподіл χ^2 («хі-квадрат») із $k = n-1$ ступенями свободи:

$$\chi^2(k) = \frac{S^2}{\sigma^2}(n-1). \qquad (3.11)$$

Двосторонні критичні межі χ_1^2 і χ_2^2 цього розподілу визначаються із умови $P(\chi_1^2 < \chi^2(k) < \chi_2^2) = \gamma$. Після підстановки рівності (3.11) і розв'язання нерівності $\chi_1^2 < \frac{S^2}{\sigma^2}(n-1) < \chi_2^2$ відносно σ^2 отримано співвідношення

$$P\left(\chi_1^2 < \frac{S^2}{\sigma^2}(n-1) < \chi_2^2\right) = P\left(\frac{(n-1)S^2}{\chi_2^2} < \sigma^2 < \frac{(n-1)S^2}{\chi_1^2}\right) = \gamma.$$

Отже, *надійний інтервал для оцінки дисперсії $D_Г$ і середнього квадратичного відхилення σ при невідомому математичному сподіванні* μ матиме вигляд:

$$\frac{(n-1)S^2}{\chi_2^2} < D_Г < \frac{(n-1)S^2}{\chi_1^2}, \qquad (3.12)$$

$$\frac{S\sqrt{n-1}}{\chi_2} < \sigma < \frac{S\sqrt{n-1}}{\chi_1}, \tag{3.13}$$

де χ_1^2 і χ_2^2 знаходяться за таблицею значень розподілу χ^2 (додаток 8) із умов:

$$P(\chi^2(k) > \chi_1^2) = 1 - \frac{\alpha}{2};\ P(\chi^2(k) > \chi_2^2) = \frac{\alpha}{2}; \tag{3.14}$$

де $\alpha = 1 - \gamma$. Межі оцінки (3.15) не симетричні відносні S.

Інтервальну оцінку для σ при заданій надійності γ можна шукати у вигляді

$$S \cdot \max\{0; 1-q\} < \sigma < S(1+q),\ q \neq 1. \tag{3.15}$$

Дійсно, розв'язавши нерівність (3.15) відносно $\frac{(n-1)S^2}{\sigma^2}$, можна отримати, що із ймовірністю γ виконується нерівність $\frac{n-1}{(1+q)^2} < \frac{(n-1)S^2}{\sigma^2} < \frac{n-1}{\max\{0;1-q\}}$,

тобто $P\left(\frac{n-1}{(1+q)^2} < \chi^2_{(n-1)} < \frac{n-1}{\max\{0;1-q\}}\right) = \gamma$.

Отже,

$$P\left(S \cdot \max\{0;1-q\} < \sigma < S(1+q)\right) = P\left(S^2 \cdot \max{}^2\{0;1-q\} < \sigma^2 < S^2(1+q)^2\right) = \gamma.$$

Значення q знаходиться за таблицею значень (додаток 5) при заданих γ і $k = n-1$: $q = q(\gamma, k)$. При $q < 1$ інтервальна оцінка (3.15) симетрична відносно S.

Надійний інтервал для оцінки дисперсії $D_Г$ *і середнього квадратичного відхилення* σ *нормального розподілу при відомому математичному сподіванні* μ.

Ефективною оцінкою дисперсії у такому випадку буде величина

$$S_0^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2.$$

Потрібно розглянути випадкову величину $\chi^2(n)$, яка має розподіл χ^2 із $k = n$ ступенями свободи:

$$\chi^2(n) = \frac{nS_0^2}{\sigma^2}. \tag{3.16}$$

Провівши для статистики (3.16) міркування, аналогічні попереднім при невідомому математичному сподіванні, можна отримати відповідні інтервальні оцінки для D_Γ і σ:

$$\frac{nS_0^2}{\chi_2^2} < D_\Gamma < \frac{nS_0^2}{\chi_1^2}, \tag{3.17}$$

$$\frac{S_0\sqrt{n}}{\chi_2} < \sigma < \frac{S_0\sqrt{n}}{\chi_1} \tag{3.18}$$

або

$$S_0 \cdot \max\{0; 1-q\} < \sigma < S_0(1+q). \tag{3.19}$$

Значення χ_1^2, χ_2^2, q знаходяться за тими ж таблицями значень, які були вказані раніше, відмінність лише у тому, що число ступенів свободи $k = n$.

Для знаходження значень розподілу χ^2:

- використовується статистична функція **ХИ2ОБР(*вероятность*; *степени_свободы*)**, яка повертає значення розподілу χ^2, де *вероятность* – відповідне значення ймовірності, *степени_свободы* – число ступенів свободи, яке характеризує розподіл χ^2;
- використовується функція **qchisk(*p*,*k*)**, де *p* – ймовірність, *k* – кількість ступенів свободи.

Приклад 3.6.

Розміри основних фондів (x_i) на 50-ти випадково відібраних підприємствах подані у вигляді інтервального статистичного розподілу:

x_i, млн. грн.	1-1,5	1,5-2	2-2,5	2,5-3
n_i	5	15	20	10

З надійністю $\gamma = 0,98$ визначити надійний інтервал для оцінки дисперсії $D_Г$.

Розв'язання.

Оскільки математичне сподівання μ невідоме, то надійний інтервал для оцінки дисперсії $D_Г$ знаходиться за формулою (3.12):

$$\frac{(n-1)S^2}{\chi_2^2} < D_Г < \frac{(n-1)S^2}{\chi_1^2}.$$

Щоб обчислити S^2 потрібно знайти $\overline{x}_в$, $D_в$ для цього потрібно перейти до дискретного статистичного розподілу:

x_i	1,25	1,75	2,25	2,75
n_i	5	15	20	10

За дискретним статистичним розподілом:

$$\overline{x}_в = \frac{1}{n}\sum_{i=1}^{k} x_i n_i = \frac{1}{50}(1,25\cdot 5 + 1,75\cdot 15 + 2,25\cdot 20 + 2,75\cdot 10) = 2,1;$$

$$D_в = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_в)^2 n_i = \frac{1}{50}((1,25-2,1)^2\cdot 5 + (1,75-2,1)^2\cdot 8 + (2,25-2,1)^2\cdot 20 +$$

$$+(2,75-2,1)^2\cdot 10) \approx 0,203;$$

$$S^2 = \frac{n}{n-1}\cdot D_в = \frac{50}{49}\cdot 0,203 \approx 0,207.$$

Значення χ_1^2, χ_2^2 знаходяться за таблицею значень розподілу χ^2 із умов (3.14) при $k = n-1 = 50-1 = 49$ та $\alpha = 1-\gamma = 1-0,98 = 0,02$:

$$P(\chi^2(k) > \chi_1^2) = 1 - \frac{0,02}{2} = 0,99; \; P(\chi^2(k) > \chi_2^2) = \frac{0,02}{2} = 0,01;$$

$$\chi_1^2(0,99;49) = 28,9; \; \chi_2^2(0,01;49) = 74,9.$$

Тоді надійний інтервал для оцінки дисперсії $D_Г$:

$$\frac{(50-1)\cdot 0,207}{74,9} < D_Г < \frac{(50-1)\cdot 0,207}{28,9} \quad \text{або} \quad 0,135 < D_Г < 0,351.$$

	A	B	C	D	E	F	G	H	I	J
1	x_i	n_i								
2	1,25	5	$\bar{x}_в$	2,1	=СУММПРОИЗВ(A2:A5;B2:B5)/B8					
3	1,75	15								
4	2,25	20	S2	0,20663	=СУММПРОИЗВ(СТЕПЕНЬ(A2:A5-D2;2);B2:B5)/(B8-1)					
5	2,75	10	=СУММ(B2:B7)							
6	**обсяг:**	**50**								
7	γ	0,98								
8	χ_1^2	28,9406	=ХИ2ПОБР(1-B9;B6-1)			D_1	0,13515	(B6-1)*D4/B10		
9										
10	χ_2^2	74,9195	=ХИ2ПОБР(1-B9;B6-1)			D_2	0,34985	=(B6-1)*D4/B8		

$$ORIGIN := 1$$

$$IN := \begin{pmatrix} 1 & 1.5 & 5 \\ 1.5 & 2 & 15 \\ 2 & 2.5 & 20 \\ 2.5 & 3 & 10 \end{pmatrix} \qquad n := \sum IN^{\langle 3 \rangle} = 50$$

$$X := \left| \begin{array}{l} \text{for } i \in 1..\,\text{rows}(IN) \\ \quad x_{i,1} \leftarrow \dfrac{IN_{i,1} + IN_{i,2}}{2} \\ x \end{array} \right. \qquad X = \begin{pmatrix} 1.25 \\ 1.75 \\ 2.25 \\ 2.75 \end{pmatrix}$$

$$xv := \frac{1}{n} \cdot \sum_{i=1}^{rows(X)} \left(X_{i,1} \cdot IN_{i,3} \right) = 2.1$$

$$Dv := \frac{1}{n} \cdot \sum_{i=1}^{rows(X)} \left[\left(X_{i,1} - xv \right)^2 \cdot IN_{i,3} \right] = 0.203 \qquad S2 := \frac{n}{n-1} \cdot Dv = 0.207$$

$$\gamma := 0.98 \qquad \alpha := 1 - \gamma$$

$$\chi 21 := qchisq\left(\frac{\alpha}{2}, n-1 \right) = 28.941 \qquad \chi 22 := qchisq\left(1 - \frac{\alpha}{2}, n-1 \right) = 74.919$$

$$D1 := \frac{(n-1)S2}{\chi 22} = 0.135 \qquad D2 := \frac{(n-1)S2}{\chi 21} = 0.35$$

Відповідь: надійний інтервал для оцінки дисперсії: $0{,}135 < D_Г < 0{,}351$.

Приклад 3.7.

Досліджувався прибуток (тис. грн.) 45 випадково відібраних малих підприємств. З надійністю $\gamma = 0{,}95$ визначити надійний інтервал для оцінки середнього квадратичного відхилення σ, якщо $\bar{x}_в = 243$, $S = 24$.

Розв'язання.

Оскільки математичне сподівання μ відоме, то надійний інтервал для оцінки середнього квадратичного відхилення σ знаходиться за формулою (3.19):

$$S_0 \cdot \max\{0; 1-q\} < \sigma < S_0(1+q).$$

Значення q знаходиться за таблицею значень (додаток 5) при заданих $\gamma = 0{,}95$ і $k = n = 45$: $q = q(0{,}95; 45) = 0{,}22$.

Тоді надійний інтервал для середнього квадратичного відхилення σ:

$$24 \cdot \max\{0; 1-0{,}22\} < \sigma < 24(1+0{,}22) \text{ або } 18{,}72 < \sigma < 29{,}28.$$

Відповідь: надійний інтервал для оцінки середнього квадратичного відхилення: $18{,}72 < \sigma < 29{,}28$.

3.6. Оцінка ймовірності появи (частки) ознаки при великому обсязі вибірки

Відомо, що оцінкою ймовірності p події є відносна частота появи цієї події $p^* = \frac{m}{n} = w.$ Ця оцінка є незсуненою, оскільки $M(p^*) = M(w) = p$ і ґрунтовною в силу закону великих чисел (теорема Бернуллі).

При великому обсязі вибірки $(n \geq 30)$ розподіл величини $p = w$ можна вважати нормальним із параметрами $M(w) = p,\quad \sigma(w) = \sqrt{\frac{pq}{n}}\quad (q = 1 - p).$ Розрахунок надійного інтервалу проводиться за формулою:

$$P(|w - p| < \delta) = 2\Phi\left(\frac{\delta}{\sigma(w)}\right) = \gamma, \tag{3.20}$$

звідки можна отримати нерівність $|p - w| < t \cdot \sqrt{\frac{pq}{n}},$ яка визначається із ймовірністю γ. Значення t знаходиться за таблицею значень функції Лапласа при умові, що $\Phi(t) = \frac{\gamma}{2}.$

Замінивши значення p і $q = 1 - p$ в останній нерівності їх оцінками w і $w - 1,$ можна отримати *надійний інтервал для ймовірності* p:

$$w - t \cdot \sqrt{\frac{w(1-w)}{n}} < p < w + t \cdot \sqrt{\frac{w(1-w)}{n}}. \tag{3.21}$$

<u>Приклад 3.8.</u>

На хлібокомбінаті проводилась перевірка якості продукції, для цього із великої партії випечених паляниць було відібрано 120, серед них 12 не відповідали нормі за вагою. Потрібно знайти:

1) 99 % надійний інтервал для ймовірності появи якісної продукції у всій партії;

2) ймовірність того, що відхилення частки якісної продукції у вибірці від ймовірності появи такої продукції у всій партії за абсолютною величиною не перевищить 3 %.

<u>Розв'язання.</u>

1) За умовою $n = 150,$ $m = 150 - 12 = 138,$ отже, частка якісної продукції у вибірці становить $w = \frac{m}{n} = \frac{138}{150} = 0,92.$

Надійний інтервал для ймовірності появи якісної продукції у всій партії визначається за формулою (3.21):

$$w - t \cdot \sqrt{\frac{w(1-w)}{n}} < p < w + t \cdot \sqrt{\frac{w(1-w)}{n}}.$$

Значення t знаходиться за таблицею значень функції Лапласа при умові, що $\Phi(t) = \frac{\gamma}{2} = \frac{0,99}{2} = 0,495,$ тоді $t = 2,58.$

Отже,

$$0,92 - 2,58 \cdot \sqrt{\frac{0,92(1-0,92)}{150}} < p < 0,92 + 2,58 \cdot \sqrt{\frac{0,92(1-0,92)}{150}}$$

$$\text{або } 0,86 < p < 0,98.$$

2) Оскільки $\delta = 0,03;$ $\sigma(w) = \sqrt{\frac{pq}{n}} \approx \sqrt{\frac{w(1-w)}{n}} = \sqrt{\frac{0,92(1-0,92)}{150}} \approx 0,022;$ то за формулою (3.20):

$$P(|w - p| < \delta) = 2\Phi\left(\frac{\delta}{\sigma(w)}\right) = \gamma \text{ або}$$

$$P(|0,92 - p| < 0,03) = 2\Phi\left(\frac{0,03}{0,022}\right) = 2\Phi(1,36) = 2 \cdot 0,4131 \approx 0,8.$$

	A	B	C	D	E	F	G	H
1	n	150						
2	k	12						
3	m=n-k	138						
4	γ	0,99						
5	w=m/n	0,92	=НОРМОБР(B4/2+0,5;0;1)					
6	t	2,5758293						
7			=B5-(B7*КОРЕНЬ(B5*(1-B5)/B1))					
8	p_1	0,86294281						
9			=B5+(B7*КОРЕНЬ(B5*(1-B5)/B1))					
10	p_2	0,97705719						
11					КОРЕНЬ(B5*(1-B5)/B1)			
12	δ	0,03	σ	0,02215				
13			=B12/D12		=2*(НОРМСТРАСП(B14)-0,5)			
14	t	1,35434085	p	0,82437				

1)

$$n := 150 \quad k := 12 \quad \gamma := 0.99$$

$$m := n - k \quad w := \frac{m}{n} \quad \alpha := 1 - \gamma$$

$$t := \mathrm{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) = 2.576$$

$$p_1 := w - t \cdot \sqrt{\frac{w \cdot (1 - w)}{n}} = 0.863 \qquad p_2 := w + t \cdot \sqrt{\frac{w \cdot (1 - w)}{n}} = 0.977$$

2)

$$\delta := 0.03$$

$$\sigma := \sqrt{\frac{w \cdot (1 - w)}{n}} = 0.022 \qquad t := \frac{\delta}{\sigma} = 1.354$$

$$P := 2 \cdot (\mathrm{pnorm}(t, 0, 1) - 0.5) = 0.824$$

Відповідь: 1) з надійністю $\gamma = 0{,}99$ генеральна частка якісної продукції в усій партії не виходить за межі інтервалу $(0{,}89; 0{,}98)$; 2) ймовірність того, що відхилення частки якісної продукції у вибірці від ймовірності появи такої продукції у всій партії за абсолютною величиною не перевищить 3 %, дорівнює $0{,}8$.

Надійні інтервали для оцінок математичного сподівання, дисперсії, середнього квадратичного відхилення та ймовірності появи ознаки, зведені у табл. 3.1.

Таблиця 3.1

Припу-щення	Надійний інтервал	Примітки
1	**2**	**3**
Інтервальні оцінки для математичного сподівання для $M(X)=\mu$		
σ - відоме	$\bar{x}_в - \frac{t\sigma}{\sqrt{n}} < \mu < \bar{x}_в + \frac{t\sigma}{\sqrt{n}}$	Значення t знаходиться за таблицею значень функції Лапласа при умові, що $\Phi(t)=\gamma/2$. Якщо обсяг вибірки n великий, то цю інтервальну оцінку можна використовувати і за відсутності нормального розподілу величини X.
σ - невідоме	$\bar{x}_в - \frac{S\cdot t_\gamma}{\sqrt{n}} < \mu < \bar{x}_в + \frac{S\cdot t_\gamma}{\sqrt{n}}$	Випадкова величина t_γ знаходиться за таблицею значень розподілу Стьюдента при заданих значеннях γ і k: $t_\gamma = t(\gamma,k)$. При $k=n-1>30$ випадкова величина t_γ має розподіл, близький до нормального, тому знаходиться за таблицею значень функції Лапласа із умови $\Phi(t)=\gamma/2$.
Інтервальні оцінки для дисперсії $D_Г=\sigma^2$ *та середнього квадратичного відхилення* σ		
μ - невідоме	$\frac{(n-1)S^2}{\chi_2^2} < D_Г < \frac{(n-1)S^2}{\chi_1^2}$, $\frac{S\sqrt{n-1}}{\chi_2} < \sigma < \frac{S\sqrt{n-1}}{\chi_1}$ або $S\cdot\max\{0;1-q\} < \sigma < S(1+q),\ q\neq 1$ (якщо задана γ)	Значення χ_1^2, χ_2^2 знаходяться із таблиці за умов: $P(\chi^2(k)>\chi_1^2)=1-\frac{\alpha}{2}$; $P(\chi^2(k)>\chi_2^2)=\frac{\alpha}{2}$; де $\alpha=1-\gamma$. Межі оцінки не симетричні відносно S. Значення q знаходиться із таблиці при заданих γ і $k=n-1$: $q=q(\gamma,k)$. При $q<1$ інтервальна оцінка симетрична відносно S.

1	2	3
μ - відоме	$\dfrac{nS_0^2}{\chi_2^2} < D_Г < \dfrac{nS_0^2}{\chi_1^2}$, $\dfrac{S_0\sqrt{n}}{\chi_2} < \sigma < \dfrac{S_0\sqrt{n}}{\chi_1}$ або $S_0 \cdot \max\{0;1-q\} < \sigma < S_0(1+q)$ (якщо задана γ)	Значення χ_1^2, χ_2^2, q визначаються за тими ж таблицями, але число ступенів свободи $k = n$.
Інтервальна оцінка для ймовірності появи (частки) ознаки		
-	$w - t \cdot \sqrt{\dfrac{w(1-w)}{n}} < p < w + t \cdot \sqrt{\dfrac{w(1-w)}{n}}$	Значення t знаходиться за таблицею значень функції Лапласа при умові, що $\Phi(t) = \gamma/2$.

Практичне заняття 3

3.1. Для оцінки якості продукції підприємства «М'яско» було проведене опитування споживачів. У декількох магазинах були розміщені скриньки, в які пропонувалось помістити одну із оцінок від 1 до 5. У результаті було отримано статистичний розподіл:

x_i	1	2	3	4	5
n_i	10	27	33	49	23

Знайти точкові незсунені статистичні оцінки генеральних середньої та дисперсії.

3.2. Знайти точкові незсунені статистичні оцінки генеральних середньої та дисперсії за дискретним статистичним розподілом вибірки:

а)

x_i	-1	3	7	11
n_i	2	3	12	8

б)

x_i	105	109	114	121
n_i	1	15	10	9

У завданнях 3.3 – 3.15 вважається, що ознака має нормальний розподіл.

3.3. На автоматизованій технологічній лінії кондитерського цеху виготовляються тістечка «Корівка». Для контролю довідним чином було відібрано 25 тістечок і виміряно їх масу, отримані результати подано у вигляді статистичного розподілу:

x_i, г	147-149	149-151	151-153
n_i	8	11	6

Потрібно: 1) знайти точкові незсунені статистичні оцінки генеральних середньої та дисперсії; 2) з надійністю $\gamma = 0,95$ визначити надійний інтервал для дійсного середнього значення маси тістечка.

3.4. У 20 філіалах концерну «Всесвіт» було проведено ряд інноваційних заходів, що призвело до зростання прибутку. Результати дослідження відсотку зростання прибутку на підприємствах подано у вигляді статистичного розподілу:

x_i, %	0,2	0,5	0,8	1,1
n_i	5	11	3	1

Потрібно: 1) знайти точкові незсунені статистичні оцінки генеральних середньої та дисперсії; 2) з надійністю $\gamma = 0,98$ визначити надійний інтервал для дійсного середнього значення відсотку зростання прибутку.

3.5. Після зважування 25, відібраних для контролю ваги, пачок майонезу було визначено вибіркову середню $\overline{x}_в = 101$ г. Із надійністю $\gamma = 0,98$

побудувати надійний інтервал для дійсного середнього значення маси пачки майонезу $\overline{X}_{Г}$, якщо середнє квадратичне відхилення $\sigma = 0{,}5$.

3.6. Дано дискретний статистичним розподіл вибірки:

а)

x_i	-5,2	3,4	8,5
n_i	3	10	5

б)

x_i	250	255	260	265
n_i	3	10	12	5

З надійністю $\gamma = 0{,}95$ визначити надійний інтервал для дійсного середнього значення.

3.7. Дано дискретний статистичним розподіл вибірки:

а)

x_i	100	120	130
n_i	2	10	3

б)

x_i	105	110	115	120
n_i	3	11	9	7

З надійністю $\gamma = 0{,}95$ визначити надійний інтервал для дійсного середнього значення, якщо $\sigma = 5$.

3.8. Якого значення має набути надійність оцінки γ, щоб за обсягу вибірки $n = 144$ і $\sigma = 1$ її похибка не перевищувала 0,1.

3.9. Якого значення має набути надійність оцінки γ, щоб за обсягу вибірки $n = 100$ і $\sigma = 1$ її похибка не перевищувала 0,01.

3.10. Визначити мінімальний обсяг вибірки $n_{\min}$, за якого похибка $\delta = 0{,}01$ гарантувалася б з ймовірністю $0{,}99$, якщо $\sigma = 2$.

3.11. Визначити мінімальний обсяг вибірки $n_{\min}$, за якого похибка $\delta = 0{,}1$ гарантувалася б з ймовірністю $0{,}98$, якщо $\sigma = 1$.

3.12. Розміри основних фондів (x_i) на 50-ти випадково відібраних підприємствах подані у вигляді інтервального статистичного розподілу:

x_i, млн. грн.	0,5-1	1-1,5	1,5-2	2-2,5
n_i	5	15	20	10

З надійністю $\gamma = 0{,}98$ визначити надійний інтервал для оцінки генеральної дисперсії.

3.13. Для перевірки ваги відібрано 30 банок консервів. Відхилення від номінальної ваги банки консервів подано у вигляді статистичного розподілу:

x_i, г	0	0,1	0,2	0,3
n_i	5	15	6	4

З надійністю $\gamma = 0{,}98$ визначити надійний інтервал для оцінки середнього квадратичного відхилення.

3.14. Досліджувалась величина прибутку від впровадження нового виду йогуртів на 10 підприємствах концерну. З надійністю $\gamma = 0{,}95$ визначити надійний інтервал для оцінки генеральної дисперсії, якщо $\overline{x}_в = 243$, $S = 24$.

3.15. На підприємстві проводилась перевірка якості продукції, для цього із великої партії пакетів соків було відібрано 130, серед них 15 мали пошкоджені пакети. Знайти:

1) 99 % надійний інтервал для ймовірності появи якісної продукції у всій партії;

2) ймовірність того, що відхилення частки якісної продукції у вибірці від ймовірності появи такої продукції у всій партії за абсолютною величиною не перевищить 1 %.

3.16. У колл-центрі інформаційного відділу досліджувалась кількість абонентів, які не дочекались з’єднання з оператором. Виявилось, що із 100 абонентів 9 переривали зв’язок. Знайти:

1) 95 % надійний інтервал для ймовірності переривання зв’язку у генеральній сукупності;

2) ймовірність того, що відхилення частки переривань зв’язку у вибірці від ймовірності переривання зв’язку у генеральній сукупності за абсолютною величиною не перевищить 2 %.

Лекція 4. Перевірка статистичних гіпотез

4.1. Статистичні гіпотези та критерії їх перевірки

4.2. Загальний алгоритм перевірки статистичної гіпотези

4.3. Перевірка достовірності гіпотези про значення генеральної середньої

4.4. Перевірка гіпотези про значення ймовірності (частки ознаки) у генеральній сукупності

4.1. Статистичні гіпотези та критерії їх перевірки

Інформація, отримана внаслідок дослідження вибірки з генеральної сукупності, може використовуватися для деяких висновків щодо всієї генеральної сукупності. Наприклад, на підприємстві про якість продукції судять за результатами вибіркового контролю. Якщо вибіркова частка браку не перевищує заздалегідь встановленої (нормативної) величини p_0, то партія продукції приймається. Оскільки висновки про відповідність якості продукції встановленим вимогам робляться на підставі вибіркової перевірки, тому судження про якість продукції не можна розглядати як категоричне. Тобто мова йде лише про припущення (гіпотезу), що частка браку в усій партії (генеральній сукупності) менша або дорівнює p_0.

<u>*Означення.*</u> *Статистичною гіпотезою* називається будь-яке судження про вигляд або параметри невідомого закону розподілу генеральної сукупності, зроблене на підставі вибірки.

Статистичні гіпотези поділяються на *параметричні* (містять твердження про значення параметрів ознак генеральної сукупності) та *непараметричні* (містять твердження про закон розподілу ознаки генеральної сукупності).

Гіпотеза, що підлягає перевірці, називається *основною* або *нульовою* і позначається H_0. Зміст нульової гіпотези записується так: $H_0 : \bar{X}_Г = \mu$ або $H_0 : p = p_0$, або $H_0 : \sigma = \sigma_0$. Кожній нульовій гіпотезі можна протиставити кілька *альтернативних* або *конкуруючих* гіпотез, які позначаються через H_1. Альтернативна гіпотеза заперечує нульову.

Розрізняються *прості* та *складні* статистичні гіпотези.

Проста параметрична гіпотеза характеризує параметр розподілу ознаки генеральної сукупності однозначно, наприклад, $H_0 : \overline{X}_Г = 100$.

Складна параметрична гіпотеза неоднозначна, вона може стверджувати, що параметр розподілу ознаки генеральної сукупності належить певній множині можливих значень, наприклад, $H_0 : 0 < \sigma < 2$.

Для перевірки відповідності даних вибірки сформульованій нульовій гіпотезі H_0 вибирається статистичний критерій. *Статистичний критерій* – випадкова величина U, точний або наближений розподіл якої заздалегідь відомий: нормальний (Z), Стьюдента (t), Фішера (F), хі-квадрат (χ^2) тощо. Керуючись статистичним критерієм нульову гіпотезу приймають або відхиляють. При великих обсягах вибірки ($n > 30$) закони розподілів статистичний критеріїв $t,\ F,\ \chi^2$ наближаються до нормального розподілу.

Множину всіх значень статистичного критерію U поділяють на дві підмножини G та $\overline{G}$: $G \cap \overline{G} = \varnothing$, $G \cup \overline{G} = \Omega$. Множина $\overline{G}$ називається *критичною областю*, вона містить значення критерію, при яких нульова гіпотеза відхиляється; G – область прийняття гіпотези H_0 (область допустимих значень). Точку або точки, які поділяють множину всіх значень критерію на критичну область і область прийняття гіпотези називають критичними і позначають $u_{кр}$.

За даними вибірки обчислюється спостережуване значення u^* статистичного критерію U. Якщо число u^* потрапляє в критичну область $\overline{G}$, то гіпотеза H_0 відхиляється і приймається альтернативна гіпотеза H_1; якщо ж значення u^* потрапляє в область допустимих значень G, то приймається H_0, а H_1 відхиляється. Яку б велику точність не мали проведені розрахунки, завжди є ймовірність відхилити гіпотезу H_0, при умові, що вона є правильною. Тому при перевірці правильності гіпотези H_0 можуть бути допущені помилки. Отже, можливі чотири випадки (табл. 4.1).

Таблиця 4.1

Гіпотеза H_0	**Рішення**	
	Приймається	**Відхиляється**
Вірна	Правильне рішення	Помилка 1-го роду
Невірна	Помилка 2-го роду	Правильне рішення

<u>*Означення.*</u> Ймовірність α припустити помилку першого роду, тобто $\alpha = P(H_1 / H_0)$, називається *рівнем значущості*.

Величина α задається наперед і за звичай приймається рівною 0,001; 0,005; 0,01; 0,05.

Ймовірність припустити помилку другого роду позначають через β: $\beta = P(H_0 / H_1)$.

<u>*Означення.*</u> Ймовірність $1 - \beta = P(H_1 / H_1)$ не припустити помилку другого роду, тобто відхилити гіпотезу H_0, коли вона невірна, називається *потужністю критерію*.

При заданому рівні значущості, потужність бажано мати найбільшою, що гарантує найменшу ймовірність помилки другого роду.

Розрізняють три види критичних областей: правостороння, лівостороння і двостороння.

Правосторонньою називається критична область, яка задається нерівністю $U > u_{кр}$, де $u_{кр} > 0$ (рис. 4.1). Критичну точку $u_{кр}$ при обраному рівні значущості визначають за співвідношенням

$$P(U > u_{кр}) = \alpha. \tag{4.1}$$

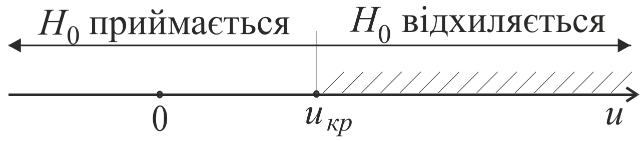


Рис 4.1

Лівосторонньою називається критична область, яка задається нерівністю $U < u_{кр}$, де $u_{кр} < 0$ (рис. 4.2). Критичну точку $u_{кр}$ при обраному рівні значущості визначають за співвідношенням

$$P(U < u_{кр}) = \alpha. \tag{4.2}$$

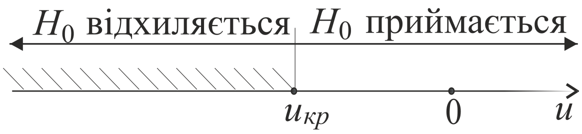


Рис 4.2

Двосторонньою називається критична область, яка задається нерівностями $U < u'_{кр}$, $U > u''_{кр}$ (рис. 4.3). Критичні точки $u'_{кр}$, $u''_{кр}$ при обраному рівні значущості визначають за умови

$$P(U < u'_{кр}) + P(U > u''_{кр}) = \alpha. \tag{4.3}$$

Якщо область симетрична відносно нуля, тобто $P(U < u'_{кр}) = P(U > u''_{кр}) = \dfrac{\alpha}{2}$ і $u''_{кр} = -u'_{кр} = u_{кр}$, то $u_{кр}$ визначається за рівнянням

$$P(U > u_{кр}) = \frac{\alpha}{2}. \tag{4.4}$$

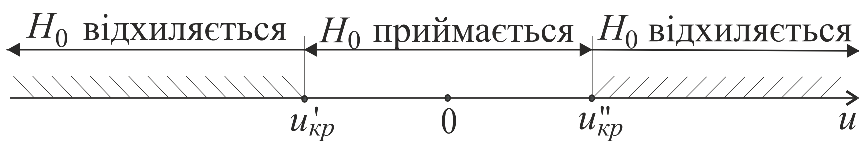


Рис. 4.3

4.2. Загальний алгоритм перевірки статистичної гіпотези

1. Формулюються гіпотези: нульова H_0 і альтернативна H_1.

2. Вибирається статистичний критерій, який відповідає сформульованій нульовій гіпотезі: $U \in \{Z, t, F, \chi^2\}$.

3. Залежно від змісту альтернативної гіпотези за вибраним статистичним критерієм U та рівнем значущості α визначаються критичні точки розподілу критерію і будується правобічна, лівобічна або двобічна критична область.

4. За даними вибірки обчислюється спостережуване значення критерію $u*$.

5. Робиться висновок. Якщо $u*$ належить критичній області, то гіпотеза H_0 відхиляється, якщо ж $u*$ належить області допустимих значень, то H_0 приймається.

Прикладні задачі перевірки статистичних гіпотез можна поділити на декілька основних типів:

1) про числові значення параметрів генеральної сукупності (математичного сподівання, дисперсії, ймовірності тощо);

2) порівняння числових характеристик двох сукупностей;

3) про закон розподілу генеральної сукупності.

Вибір конкретного критерію перевірки в кожному із наведених випадків залежить від обсягів вибірки (великий чи малий) і наявності додаткової інформації щодо ознаки, яка досліджується.

4.3. Перевірка достовірності гіпотези про значення генеральної середньої

При перевірці достовірності гіпотези про значення генеральної середньої потрібно розглянути два випадки, коли генеральна дисперсія відома і коли вона невідома.

Генеральна дисперсія відома.

Нехай із генеральної сукупності взято вибірку обсягом n і обчислено $\overline{x}_в$. При рівні значущості α потрібно перевірити достовірність статистичної гіпотези $H_0 : M(X) = \mu,$ тобто генеральна середня дорівнює деякому числу $\mu.$ Такі задачі зустрічаються при вирішенні технологічних та економічних проблем, які характеризуються деяким середнім показником: середній вміст нітратів у продукції, середня вологість сировини, середній обсяг реалізації продукції, середній прибуток підприємства тощо. Якщо обсяг вибірки великий $(n > 30)$ і відома генеральна дисперсія $D(X) = \sigma^2,$ то вибирається статистичний критерій

$$Z = \frac{\overline{x}_в - \mu}{\sigma(\overline{x}_в)} = \frac{(\overline{x}_в - \mu)\sqrt{n}}{\sigma}. \tag{4.5}$$

Випадкова величина Z має нормований закон розподілу ймовірностей: $Z \sim N(0;1),$ оскільки $M(Z) = 0$ і $\sigma(Z) = 1.$

При перевірці гіпотези H_0 можливі три випадки:

I. Якщо альтернативна гіпотеза має вигляд $H_1 : M(X) > \mu,$ то будується *правостороння критична область*.

Оскільки розподіл величини Z симетричний відносно нуля (рис. 4.4), то

$$P(0 < Z < \infty) = \frac{1}{2} \text{ або } P(0 < Z \le z_{кр}) + P(Z > z_{кр}) = \frac{1}{2}. \tag{4.6}$$

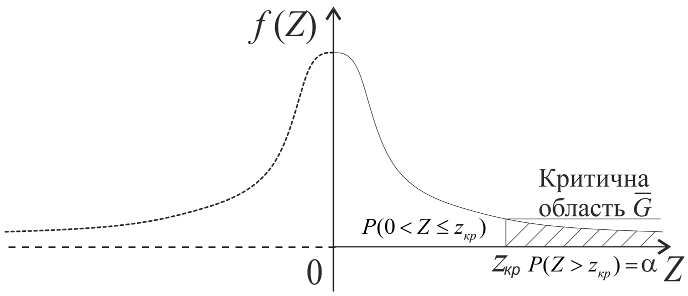


Рис. 4.4

Враховуючи те, що для правосторонньої критичної області $z_{кр}$ знаходиться із співвідношення (4.1), то $P(0 < Z \le z_{кр}) + \alpha = \frac{1}{2},$ тобто

$$P(0 < Z \le z_{кр}) = \frac{1-2\alpha}{2}. \tag{4.7}$$

Ймовірність потрапляння нормально розподіленої випадкової величини на проміжок $(0; z_{кр})$ знаходиться за формулою $\Phi(z_{кр}) - \Phi(0) = \frac{1-2\alpha}{2},$ де $\Phi(x)$ – функція Лапласа. Оскільки $\Phi(0) = 0,$ то

$$\Phi(z_{кр}) = \frac{1-2\alpha}{2}. \tag{4.8}$$

Аргумент $z_{кр}$ знаходиться за таблицею значень функції Лапласа із умови (4.8).

За наявною вибіркою обчислюється спостережуване значення критерію

$$z^* = \frac{\sqrt{n}(\overline{x}_в - \mu)}{\sigma}. \tag{4.9}$$

Якщо $z^* < z_{кр},$ то гіпотеза H_0 приймається. У протилежному випадку H_0 відхиляється на користь гіпотези $H_1.$

II. Якщо альтернативна гіпотеза має вигляд $H_1 : M(X) < \mu,$ то будується *лівостороння критична область*, $z_{кр}$ знаходиться із співвідношення (4.2).

Враховуючи те, що виконується рівність $P(Z < z_{кр}) + P(z_{кр} \le Z < 0) = \frac{1}{2},$ можна отримати $P(z_{кр} \le Z < 0) = \frac{1-2\alpha}{2}$ або $\Phi(0) - \Phi(z_{кр}) = \frac{1-2\alpha}{2}.$ Оскільки $\Phi(0) = 0,$ то $-\Phi(z_{кр}) = \frac{1-2\alpha}{2},$ остаточно

$$\Phi(z_{кр}) = -\frac{1-2\alpha}{2}. \tag{4.10}$$

Аргумент $z_{кр}$ знаходиться за таблицею значень функції Лапласа із умови (4.10), враховуючи те, що ця функція непарна (спочатку потрібно знайти $z'_{кр}$ із умови $\Phi(z'_{кр}) = \frac{1-2\alpha}{2}$, а $z_{кр} = -z'_{кр}$). Лівостороння критична область зображена на рис. 4.5.

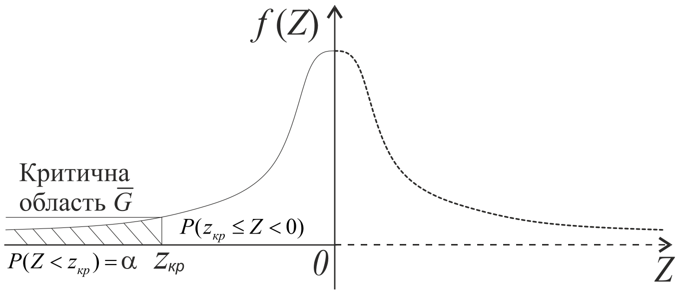


Рис. 4.5

III. Якщо альтернативна гіпотеза має вигляд $H_1 : M(X) \neq \mu$, то будується *двостороння критична область*, критичні точки $z'_{кр}$, $z''_{кр}$ $(z'_{кр} < z''_{кр})$ знаходяться із умови $P(Z < z'_{кр}) = P(Z > z''_{кр}) = \frac{\alpha}{2}$, де $z''_{кр} = -z'_{кр} = z_{кр}$.

Отже, потрібно обчислити лише значення $z_{кр}$. Скориставшись рівнянням $P(0 < Z \le z_{кр}) + P(Z > z_{кр}) = \frac{1}{2}$ та врахувавши співвідношення (4.4), можна отримати $P(0 < Z \le z_{кр}) = \frac{1-\alpha}{2}$ або $\Phi(z_{кр}) - \Phi(0) = \frac{1-\alpha}{2}$. Оскільки $\Phi(0) = 0$, то

$$\Phi(z_{кр}) = \frac{1-\alpha}{2}. \qquad (4.11)$$

Аргумент $z_{кр}$ знаходиться за таблицею значень функції Лапласа із умови (4.11).

Двостороння критична область зображена на рис. 4.6.

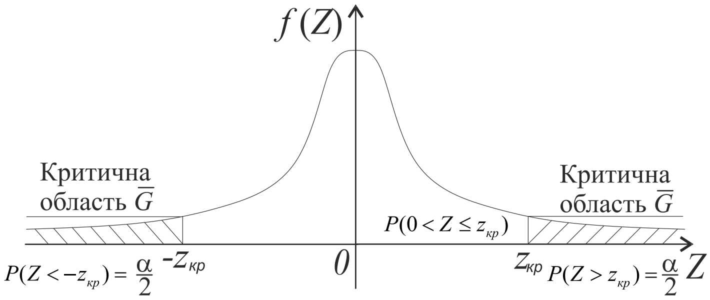


Рис. 4.6

Генеральна дисперсія невідома.

Якщо σ^2 невідома, то за статистичний критерій вибирається випадкова величина

$$t = \frac{\overline{x}_в - \mu}{S}\sqrt{n}, \tag{4.12}$$

яка має розподіл Стьюдента з $k = n - 1$ ступенями волі, $S = \sqrt{\frac{n}{n-1}D_в}$ – виправлене стандартне відхилення.

Критична область будується в залежності від вигляду альтернативної гіпотези H_1, аналогічно розглянутим випадкам з відомою дисперсією, але критичні точки $t_{кр}$ знаходяться за таблицею значень розподілу Стьюдента за відомими параметрами α і $k = n - 1$: $t_{кр} = t(\alpha, k)$.

<u>*Зауваження.*</u> При великих обсягах вибірки $(n > 30)$ випадкова величина t (4.12) має розподіл, близький до нормального, тому у такому випадку критичні точки визначаються за рівностями (4.8, 4.10, 4.11).

<u>Приклад 4.1.</u>

На підприємстві працює лінія по розливу соків у пакети по 1 л (із стандартним відхиленням 0,02 л). Для перевірки навмання відібрано 49 пакетів соку, їхня середня маса становила 0,99 л. Чи є підстави вважати, що лінія працює з відхиленнями при рівні значущості $\alpha = 0,01$?

Розв'язання.

Нехай випадкова величина X – об'єм пакету сока. З умови задачі випливає, що потрібно перевірити нульову гіпотезу $H_0: M(X)=1$ за альтернативної $H_1: M(X)\neq 1.$ Оскільки $n=49>30,$ то можна вважати, що випадкова величина X розподілена нормально.

За умовою відомо, що середнє квадратичне відхилення $\sigma=0,02,$ отже, потрібно скористатись критерієм (4.5): $Z=\dfrac{(\overline{x}_в-\mu)\sqrt{n}}{\sigma}.$ За змістом альтернативної гіпотези потрібно будувати двосторонню критичну область.

Для визначення критичного значення $z_{кр}$ слід використати формулу (4.11): $\Phi(z_{кр})=\dfrac{1-\alpha}{2},$ потім скористатись таблицею значень функції Лапласа:

$$\Phi(z_{кр})=\frac{1-\alpha}{2}=\frac{1-0,01}{2}=0,495;\ z_{кр}=2,58.$$

Спостережуване значення критерію обчислюється за вибраним критерієм (формулою (4.5)) при відомих $\overline{x}_в=0,99;\ \mu=1;\ \sigma=0,02;\ n=49:$

$$z^*=\frac{(0,99-1)\sqrt{49}}{0,02}=-3,5.$$

Двостороння критична область зображена на рис. 4.7:

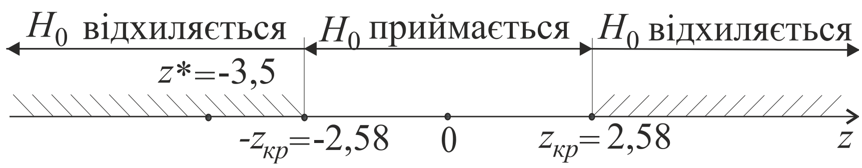


Рис. 4.7

Оскільки $z^*<-z_{кр}$ належить критичній області (рис. 4.7), то гіпотеза H_0 відхиляється, приймається альтернативна гіпотеза $H_1.$

	A	B	C	D	E
1	$\bar{x}_в$	0,99			
2	μ	1			
3	n	49			
4	σ	0,02			
5	α	0,01			
6	$z_{кр}$	2,575829			
7					
8	z	-3,5			

=НОРМОБР(1-B5/2;0;1)

=(B1-B2)*КОРЕНЬ(B3)/B4

$$n := 49 \quad \mu := 1 \quad xv := 0.99 \quad \sigma := 0.02 \quad \alpha := 0.01$$

$$zk := \mathrm{qnorm}\left(1 - \frac{\alpha}{2}, 0, 1\right) = 2.576$$

$$z := \frac{(xv - \mu)\cdot\sqrt{n}}{\sigma} = -3.5$$

<u>Відповідь:</u> при рівні значущості 0,01 можна зробити висновок, що лінія по розливу соків працює з відхиленнями, тобто її потрібно налагоджувати.

4.4. Перевірка гіпотези про значення ймовірності (частки ознаки) у генеральній сукупності

Нехай задано генеральну сукупність з ознакою X. Із генеральної сукупності взято вибірку обсягом n. При заданому рівні значущості α необхідно перевірити достовірність гіпотези $H_0 : p = p_0$ – ймовірність p появи ознаки X в генеральній сукупності дорівнює заданій величині p_0.

Потрібно розглянути випадкову величину $w = \frac{m}{n}$, яка має біномний розподіл. При великих обсягах вибірки $(n > 30)$ можна наближено вважати нормальним закон розподілу величини w з математичним сподіванням

$M(w) = p_0$, дисперсією $D(w) = \frac{p_0 q_0}{n}$, середнім квадратичним відхиленням $\sigma(w) = \sqrt{\frac{p_0 q_0}{n}}$, де $q_0 = 1 - p_0$.

В такому випадку за статистичний критерій потрібно прийняти нормовану нормально розподілену випадкову величину $Z \sim N(0;1)$:

$$Z = \frac{w - p_0}{\sigma(w)} = \frac{(w - p_0)\sqrt{n}}{\sqrt{p_0 q_0}}. \quad (4.13)$$

Критична область будується в залежності від змісту альтернативної гіпотези H_1 аналогічно, як при перевірці гіпотези про значення генеральної середньої.

<u>Приклад 4.2.</u>

На підприємстві працює інспекція по контролю якості продукції. Для вдалого інспектування частка неякісної продукції має бути меншою, ніж 1 %. Серед навмання відібраних 500 зразків продукції 10 визнані неякісними. Чи є підстави вважати за рівня значущості $\alpha = 0,05$, що на підприємстві виробляється неякісна продукція?

<u>Розв'язання.</u>

Потрібно перевірити нульову гіпотезу $H_0 : p = 0,01$ (частка неякісної продукції становить 1 %) за альтернативної $H_1 : p > 0,01$ (зростання частки неякісної продукції).

За змістом альтернативної гіпотези потрібно будувати правосторонню критичну область.

Для визначення критичного значення $z_{кр}$ слід використати формулу (4.8): $\Phi(z_{кр}) = \frac{1 - 2\alpha}{2}$, потім скористатись таблицею значень функції Лапласа:

$$\Phi(z_{кр}) = \frac{1 - 2\alpha}{2} = \frac{1 - 0,1}{2} = 0,45;\ z_{кр} = 1,645.$$

При великих обсягах вибірки $(n = 500 > 30)$ можна наближено вважати нормальним закон розподілу випадкової величини $w = \frac{m}{n} = \frac{10}{500} = 0{,}02$. Тоді спостережуване значення критерію обчислюється за критерієм (4.13) $Z = \frac{(w - p_0)\sqrt{n}}{\sqrt{p_0 q_0}}$ при відомих $w = 0{,}02$; $p_0 = 0{,}01$; $q_0 = 1 - 0{,}01 = 0{,}99$; $n = 500$:

$$z^* = \frac{(0{,}02 - 0{,}01)\sqrt{500}}{\sqrt{0{,}01 \cdot 0{,}99}} \approx 2{,}25.$$

Правостороння критична область зображена на рис. 4.8:

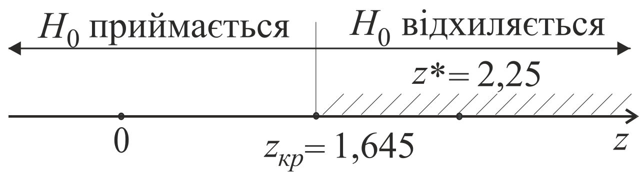


Рис. 4.8

Оскільки $z^* > z_{кр}$ належить критичній області, то гіпотеза H_0 відхиляється, приймається альтернативна гіпотеза H_1.

<u>Відповідь:</u> при рівні значущості 0,05 можна зробити висновок, що на підприємстві виробляється неякісна продукція.

Критерії перевірки гіпотез про числові значення параметрів нормального закону розподілу зведені у табл. 4.2.

Таблиця 4.2

Гіпотеза H_0	**Припу-щення**	**Статис-тичний критерій**	**Гіпотеза** H_1	**Критична точка**	**Область прийняття гіпотези** H_0
1	**2**	**3**	**4**	**5**	**6**
$M(X) = \mu$	σ - відоме	$Z = \frac{\overline{x}_в - \mu}{\sigma/\sqrt{n}}$	$M(X) > \mu$ $M(X) < \mu$ $M(X) \neq \mu$	$\Phi(z_{кр}) = \frac{1-2\alpha}{2}$ $\Phi(z_{кр}) = -\frac{1-2\alpha}{2}$ $\Phi(z_{кр}) = \frac{1-2\alpha}{2}$	$Z < z_{кр}$ $Z > -z_{кр}$ $\lvert Z \rvert < z_{кр}$

Продовження табл. 4.2

1	2	3	4	5	6
$M(X)=\mu$	σ - невідоме	$t=\dfrac{\overline{x}_в-\mu}{S/\sqrt{n}}$	$M(X)>\mu$ $M(X)<\mu$ $M(X)\neq\mu$	$t_{кр}=t(\alpha,n-1)$ $t_{кр}=-t(\alpha,n-1)$ $t_{кр}=t\left(\dfrac{\alpha}{2},n-1\right)$	$t<t_{кр}$ $t>-t_{кр}$ $\lvert t\rvert<t_{кр}$
$p=p_0$	$n>30$ $np_0>5$ $nq_0>5$	$Z=\dfrac{w-p_0}{\sqrt{p_0q_0/n}}$, $w=\dfrac{m}{n}$, $q_0=1-p_0$	$p>p_0$ $p<p_0$ $p\neq p_0$	$\Phi(z_{кр})=\dfrac{1-2\alpha}{2}$ $\Phi(z_{кр})=-\dfrac{1-2\alpha}{2}$ $\Phi(z_{кр})=\dfrac{1-\alpha}{2}$	$Z<z_{кр}$ $Z>-z_{кр}$ $\lvert Z\rvert<z_{кр}$

Практичне заняття 4

4.1. Відомо, що для виготовлення йогуртів потрібно підтримувати температуру на рівні 40^0 С. Для контролю процесу було проведено 25 замірів температури (x_i):

x_i	37	39	40	41
n_i	4	7	9	5

При рівні значущості $\alpha=0,05$ перевірити гіпотезу $H_0:M(X)=40,$ за альтернативної гіпотези $H_1:M(X)>40.$

4.2. Підприємство випускає морозиво «Карамелька» вагою 125 г (із стандартним відхиленням 2 г). У результаті перевірки 55 пачок морозива було встановлено, що середня вага складає 124 г. При рівні значущості $\alpha=0,01$ з’ясувати, чи є підстави для переналагоджування лінія з виробництва морозива «Карамелька»?

4.3. Проведено 30 незалежних вимірювань випадкової величини X, яка розподілена за нормальним законом розподілу із $\sigma = 5$:

x_i	100	115	130	145	160
n_i	3	5	12	8	2

При рівні значущості $\alpha = 0,01$ перевірити гіпотезу $H_0 : M(X) = 140,$ за альтернативної гіпотези $H_1 : M(X) < 140.$

4.4. Виробник стверджує, що вага пачки чаю «Ранок» становить 150 г. Контролером було відібрано і зважено 15 пачок чаю: 148, 150, 151, 149, 149, 150, 147, 149, 147, 147, 150, 149, 147, 151, 150. Чи правий виробник? Припускається, що вага пачок чаю розподілена за нормальним законом. Рівень значущості $\alpha = 0,05.$

4.5. Маркетолог мережі магазинів «Великий кошик» запропонував відкрити магазин у новому регіоні. На практиці встановлено, що магазин прибутковий, якщо покупець в середньому за один візит витрачає на покупки 10 ум. од. Вибіркове опитування 50 мешканців обраного регіону виявило, що середні витрати складають 12 ум. од. за один візит у магазин, виправлене середнє квадратичне відхилення $S = 5.$ При рівні значущості $\alpha = 0,01$ з'ясувати, чи буде магазин працювати прибутково?

4.6. Головний бухгалтер підприємства стверджує, що їхній середній прибуток за рік складає 450 тис. ум. од. Аудитор робить випадкову вибірку із 64 звітів, по якій виявляє, що середній прибуток становить 465 тис. ум. од. (середнє квадратичне відхилення 10 тис. ум. од.). Чи може виявитись, що правий аудитор? Рівень значущості $\alpha = 0,05.$

4.7. Підприємство фасує цукор у пачки по 950 г. Для контролю було відібрано і зважено 20 пачок цукру: 949, 950, 951, 949, 949, 950, 947, 949, 947, 947, 950, 949, 947, 951, 948, 947, 949, 947, 951, 950. Чи є підстави вважати, що маса пачки цукру не відповідає заявленій? Припускається, що вага пачок цукру розподілена за нормальним законом. Рівень значущості $\alpha = 0,05$.

4.8. Менеджер відділу збуту підприємства постійно відслідковує зміни закупівельних можливостей постійних клієнтів. З цією метою він аналізує суми рахунків постійних клієнтів протягом кварталу. У минулому кварталі середня сума рахунків становила 1250 ум. од. У поточному кварталі, проаналізувавши 45 випадковим чином вибраних рахунків, менеджером було встановлена середня сума рахунків у розмірі 1200 ум. од. (середнє квадратичне відхилення 15 тис. ум. од.). Чи правильним буде твердження при рівні значущості $\alpha = 0,01$, що закупівельні можливості постійних клієнтів зменшились? Припустити, що сума рахунків змінюється за нормальним законом розподілу.

4.9. Проведено 70 незалежних вимірювань випадкової величини X, яка розподілена за нормальним законом розподілу із $\sigma = 0,01$:

x_i	0,10	0,15	0,20	0,25	0,30
n_i	7	15	26	17	5

При рівні значущості $\alpha = 0,05$ перевірити гіпотезу $H_0 : M(X) = 0,23$, за альтернативної гіпотези $H_1 : M(X) < 0,23$.

4.10. Встановлено, що середня заробітна плата працівників харчової промисловості у деякому регіоні в минулому році становила 3116 грн. У поточному році, проаналізувавши заробітні плати 75 випадковим чином відібраних працівників, було встановлено, що середня заробітна плата 3257 грн. (середнє квадратичне відхилення 50 грн.). Чи правильним буде твердження при рівні значущості $\alpha = 0,05$, що заробітна плата працівників харчової промисловості зросла? Припустити, що заробітна плата змінюється за нормальним законом розподілу.

4.11. За результатами 27 вимірів встановлено, що середня температура, при якій тривав технологічний процес, становила 96^0 С із середнім квадратичним відхиленням 5^0 С. За нормою технологічний процес має тривати при температурі 100^0 С. При рівні значущості $\alpha = 0,01$ перевірити гіпотезу $H_0 : M(X) = 100,$ за альтернативної гіпотези $H_1 : M(X) \neq 100.$ Припускається, що температура розподілена за нормальним законом.

4.12. Відомо, що допустима швидкість руху спеціалізованого транспорту на території підприємства становить 10 км/год. Результати вимірювань швидкості руху протягом дня подані у вигляді таблиці:

x_i	7	9	11	12
n_i	7	12	5	1

При рівні значущості $\alpha = 0,05$ з'ясувати, чи відповідала швидкість руху спеціалізованого транспорту на території підприємства допустимій?

4.13. Відомо, що близько 80 % студентів другого курсу здають екзамен із дисципліни «Теорії ймовірності та математична статистика» з першого разу. За результатами поточного року відомо, що із 50 випадково відібраних студентів вказану дисципліну здали з першого разу 37. Чи є підстави вважати, що результати поточного року не суперечать попереднім дослідженням? Рівень значущості $\alpha = 0,05.$

4.14. Страховий агент компанії «Страх» стверджує, що він надає страхові послуги в середньому 40 % потенційних клієнтів. У минулому місяці із 120 потенційних клієнтів агент надав послуги 45. Чи можна вважати твердження страхового агента вірним при рівні значущості $\alpha = 0,01.$

4.15. На підприємстві працює інспекція по контролю якості продукції. Для вдалого інспектування частка неякісної продукції має бути меншою, ніж 2 %. Серед навмання відібраних 300 зразків продукції 20 визнані неякісними. Чи є підстави вважати за рівня значущості $\alpha = 0,05,$ що на підприємстві виробляється неякісна продукція?

Лекція 5. Перевірка статистичних гіпотез (продовження)

5.1. Перевірка статистичної гіпотези про рівність генеральних середніх двох сукупностей

5.2. Перевірка статистичної гіпотези про рівність двох дисперсій

5.3. Перевірка статистичної гіпотези про рівність часток ознаки у двох сукупностях

5.4. Перевірка непараметричних статистичних гіпотез. Критерій узгодженості

5.1. Перевірка статистичної гіпотези про рівність генеральних середніх двох сукупностей

Нехай задано дві генеральні сукупності з ознаками X і Y. Ознаки мають нормальний закон розподілу і є незалежними одна від одної. З кожної сукупності взято незалежні вибірки обсягами n_x і n_y відповідно, обчислено $\overline{x}_в$ і $\overline{y}_в$.

Потрібно за вибірковим середнім при обраному рівні значущості перевірити достовірність гіпотези $H_0 : M(X) = M(Y)$. За статистичний критерій потрібно взяти випадкову величину:

$$U = \frac{\overline{x}_в - \overline{y}_в}{\sigma(\overline{x}_в - \overline{y}_в)}. \tag{5.1}$$

При цьому можливі випадки:

I. *Відомі значення дисперсій ознак генеральних сукупностей*: $D(X) = D_x$, $D(Y) = D_y$.

Оскільки для незалежних вибірок

$D(\overline{x}_в - \overline{y}_в) = D(\overline{x}_в) + D(\overline{y}_в) = \dfrac{D(X)}{n_x} + \dfrac{D(Y)}{n_y}$, то критерій (5.1) набуває вигляду:

$$Z = \frac{\overline{x}_в - \overline{y}_в}{\sqrt{\frac{D_x}{n_x} + \frac{D_y}{n_y}}}, \qquad (5.2)$$

причому $M(Z) = 0,\ \sigma(Z) = 1.$

Залежно від формулювання альтернативної гіпотези H_1 будуються відповідно правостороння, лівостороння або двостороння критичні області. Критичні точки знаходяться за таблицею значень функції Лапласа при заданому рівні значущості α.

<u>*Зауваження.*</u> Статистичним критерієм (5.2) можна користуватись при перевірці гіпотези H_0 і тоді, коли немає підстав вважати, що випадкові величини X і Y мають нормальний закон розподілу, але у цьому випадку обсяги вибірок повинні бути великими $(n_x > 30, n_y > 30)$.

II. *Значення дисперсій ознак генеральних сукупностей невідомі*, але *рівні між собою*, тобто $D(X) = D(Y) = \sigma^2$.

Тоді

$$\sigma(\overline{x}_в - \overline{y}_в) = \sigma\sqrt{\frac{1}{n_x} + \frac{1}{n_y}}. \qquad (5.3)$$

Невідому дисперсію σ^2 генеральної сукупності потрібно замінити на виправлену вибіркову оцінку S^2. Краща оцінка дисперсії досягається шляхом додавання двох вибіркових дисперсій $\sigma^2_{вx}$ і $\sigma^2_{вy}$:

$$S^2 \approx \frac{n_x\sigma^2_{вx} + n_y\sigma^2_{вy}}{n_x + n_y - 2} = \frac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}. \qquad (5.4)$$

Враховуючи (5.1), (5.3), (5.4) отримано статистичний критерій:

$$t = \frac{\overline{x}_в - \overline{y}_в}{\sqrt{n_x\sigma^2_{вx} + n_y\sigma^2_{вy}}} \cdot \sqrt{\frac{n_x n_y (n_x + n_y - 2)}{n_x + n_y}}. \qquad (5.5)$$

Випадкова величина (5.6) при малих обсягах вибірок має розподіл Стьюдента з $k = n_x + n_y - 2$ ступенями свободи.

Перевірка достовірності нульової гіпотези H_0 проводиться за загальним алгоритмом (лекція 4, п. 4.2), критичне значення знаходиться за таблицею значень розподілу Стьюдента.

<u>*Зауваження.*</u> При великих обсягах вибірок $(n_x > 30, n_y > 30)$ випадкова величина t має розподіл, близький до нормального, тому у такому випадку критичні точки знаходяться за таблицею значень функції Лапласа.

III. *Значення дисперсій ознак генеральних сукупностей* $D(X)$, $D(Y)$ *невідомі і різні, обсяги вибірок великі* $(n_x > 30, n_y > 30)$.

У цьому випадку в формулі (5.2) генеральні дисперсії можна замінити відповідними виправленими вибірковими дисперсіями S_x^2, S_y^2. Одержана при цьому випадкова величина

$$Z = \frac{\overline{x}_в - \overline{y}_в}{\sqrt{\dfrac{S_x^2}{n_x} + \dfrac{S_y^2}{n_y}}}, \tag{5.6}$$

має нормований нормальний закон розподілу і обирається за статистичний критерій.

<u>*Зауваження.*</u> При виборі належного статистичного критерію перевірки нульової гіпотези H_0 за умови, що генеральні дисперсії невідомі, попередньо потрібно встановити, чи рівні між собою генеральні дисперсії. Це питання буде розглянуте у п. 5.2.

<u>**Приклад 5.1.**</u>

На молочному комбінаті працює лінії по виробництву йогуртів. До і після її вдосконалення було проведено контроль середньої ваги стаканів йогурту. Для цього було відібрано вибірки і отримані такі результати: до вдосконалення – $n_x = 40$, $\overline{x}_в = 101$ (г), $D_x = 1{,}4$; після вдосконалення – $n_y = 45$, $\overline{y}_в = 100{,}4$ (г),

$D_y = 1,2.$ Чи є підстави вважати за рівня значущості 0,05, що середня вага стаканів йогурту після вдосконалення змінилась?

Розв’язання.

Нехай випадкова величина X – вага стаканів йогурту до вдосконалення лінії, Y – вага після вдосконалення.

Нульова гіпотеза $H_0 : M(X) = M(Y)$ (середня вага стаканів йогурту не змінилась), альтернативна – $H_1 : M(X) \neq M(Y)$ (середня вага стаканів йогурту змінилась).

За змістом альтернативної гіпотези потрібно будувати двосторонню критичну область.

Для визначення критичного значення $z_{кр}$ слід використати формулу (4.11): $\Phi(z_{кр}) = \dfrac{1-\alpha}{2}$, потім скористатись таблицею значень функції Лапласа:

$$\Phi(z_{кр}) = \frac{1-0,05}{2} = 0,475; \; z_{кр} = 1,96.$$

Оскільки дисперсії відомі та обсяги вибірок $n_x = 40 > 30,$ $n_y = 45 > 30,$ то для визначення спостережуваного значення критерію потрібно використати критерій (5.2) $Z = \dfrac{\bar{x}_в - \bar{y}_в}{\sqrt{\dfrac{D_x}{n_x} + \dfrac{D_y}{n_y}}}$: $z^* = \dfrac{101 - 100,4}{\sqrt{\dfrac{1,4}{40} + \dfrac{1,2}{45}}} \approx 2,42.$

Двостороння критична область зображена на рис. 5.1:

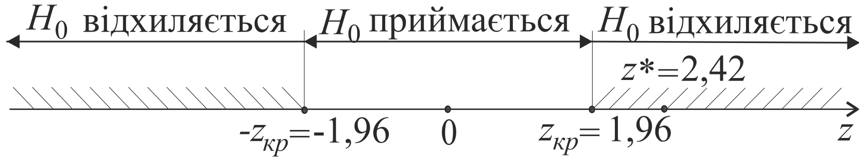


Рис. 5.1

Оскільки $z^* > z_{кр}$ належить критичній області, то гіпотеза H_0 відхиляється, приймається альтернативна гіпотеза H_1.

$nx := 40 \qquad ny := 45$

$xv := 101 \qquad yv := 100.4$

$Dx := 1.4 \qquad Dy := 1.2$

$\alpha := 0.05$

$$zk := qnorm\left(1 - \frac{\alpha}{2}, 0, 1\right) = 1.96$$

$$z := \frac{xv - yv}{\sqrt{\frac{Dx}{nx} + \frac{Dy}{ny}}} = 2.416$$

Відповідь: при рівні значущості 0,05 є підстави вважати, що середня вага стаканів йогурту після вдосконалення змінилась.

5.2. Перевірка статистичної гіпотези про рівність двох дисперсій

Однією із важливих задач математичної статистики є перевірка статистичної гіпотези про рівність двох дисперсій, тому що дисперсія характеризує такі важливі показники, як точність машин, приладів, технологічних процесів, ризик інвестицій тощо.

Нехай є дві нормально розподілені генеральні сукупності з ознаками X і Y, дисперсії яких дорівнюють σ_x^2 і σ_y^2. Необхідно перевірити нульову гіпотезу про рівність дисперсій $H_0 : \sigma_x^2 = \sigma_y^2$ за альтернативної гіпотези $H_1 : \sigma_x^2 \neq \sigma_y^2$ або $H_1 : \sigma_x^2 > \sigma_y^2$.

Для перевірки нульової гіпотези H_0 з сукупностей зроблені дві незалежні вибірки обсягами n_x і n_y, для яких знайдені виправлені вибіркові дисперсії S_x^2, S_y^2. Більшу з них позначимо через S_1^2, меншу – через S_2^2.

За статистичний критерій береться випадкова величина

$$F = \frac{S_1^2}{S_2^2}, \tag{5.7}$$

яка має розподіл Фішера-Снедекора з $k_1 = n_1 - 1,\ k_2 = n_2 - 1$ ступенями свободи.

В залежності від вигляду альтернативної гіпотези H_1 будується критична область. Критична точка $f_{кр}$ знаходиться із таблиці значень розподілу Фішера-Снедекора (додаток 7) при заданому рівні значущості α за умови: $f_{кр} = F(\alpha, k_1, k_2)$, якщо будується правостороння критична область, та за умови: $f_{кр} = F\left(\frac{\alpha}{2}, k_1, k_2\right)$, якщо будується двостороння критична область.

Для знаходження значень розподілу Фішера-Снедекора:

використовується функція **qF(*p*,*k₁*,*k₂*)**, де *p* – ймовірність, ***k₁***, ***k₂*** – кількість ступенів свободи.

Приклад 5.2.

На підприємстві працюють два цехи по виробництву олії. Відділом контролю була проведена перевірка на вміст домішок у олії. Результати дослідження подано у вигляді двох розподілів:

x_i, %	0,5	0,7	0,9	1,2
n_{x_i}	4	12	18	1

y_i, %	0,3	0,6	0,9	1,2
n_{y_i}	3	17	10	2

Ознаки X, Y (вміст домішок) є незалежними нормально розподіленими випадковими величинами. При рівні значущості $\alpha = 0,01$ перевірити гіпотезу $H_0 : D_x = D_y$ за альтернативної $H_1 : D_x > D_y$.

Розв’язання.

За змістом альтернативної гіпотези потрібно будувати правосторонню критичну область.

Для проведення подальших обчислень потрібно знайти деякі числові характеристики вибірок:

$n_x = 4+12+18+1=35;\ n_y = 3+17+10+2=32;$

$$\overline{x}_в = \frac{1}{35}(0,5\cdot 4+0,7\cdot 12+0,9\cdot 18+1,2\cdot 1)\approx 0,79;$$

$$\overline{y}_в = \frac{1}{32}(0,3\cdot 3+0,6\cdot 17+0,9\cdot 10+1,2\cdot 2)\approx 0,7;$$

$$D_{вx} = \frac{1}{35}((0,5-0,79)^2\cdot 4+(0,7-0,79)^2\cdot 12+(0,9-0,79)^2\cdot 18+$$

$$+(1,2-0,79)^2\cdot 1)\approx 0,023;$$

$$S_x^2 = \frac{35}{34}\cdot 0,023\approx 0,024;$$

$$D_{вy} = \frac{1}{32}((0,3-0,7)^2\cdot 3+(0,6-0,7)^2\cdot 17+(0,9-0,7)^2\cdot 10+(1,2-0,7)^2\cdot 2)\approx 0,048;$$

$$S_y^2 = \frac{32}{31}\cdot 0,048\approx 0,05.$$

Для визначення спостережуваного значення критерію потрібно використати критерій (5.7) $F=\dfrac{S_1^2}{S_2^2},\ (S_1^2 > S_2^2),$ де $S_1^2 = S_y^2,$ $S_2^2 = S_x^2$:

$$f* = \frac{0,05}{0,024}\approx 2,08.$$

Критичне значення $f_{кр}$ знаходиться із таблиці значень розподілу Фішера-Снедекора при заданому рівні значущості $\alpha = 0,01$ за умови $f_{кр} = F(\alpha, k_1, k_2),$ де $k_1 = n_1 - 1 = n_y - 1 = 31,\ k_2 = n_2 - 1 = n_x - 1 = 34$:

$$f_{кр} = F(0,01;31;34) = 2,2.$$

Правостороння критична область зображена на рис. 5.2:

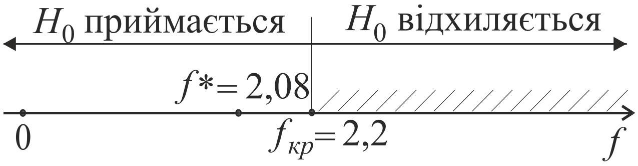


Рис. 5.2

Оскільки $f^* > f_{кр}$ не належить критичній області, то приймається гіпотеза H_0.

$$ORIGIN := 1$$

$$XN := \begin{pmatrix} 0.5 & 4 \\ 0.7 & 12 \\ 0.9 & 18 \\ 1.2 & 1 \end{pmatrix} \qquad YN := \begin{pmatrix} 0.3 & 3 \\ 0.6 & 17 \\ 0.9 & 10 \\ 1.2 & 2 \end{pmatrix}$$

$$nx := \sum XN^{\langle 2 \rangle} = 35 \qquad ny := \sum YN^{\langle 2 \rangle} = 32$$

$$xv := \frac{1}{nx} \cdot \sum_{i=1}^{rows(XN)} \left(XN_{i,1} \cdot XN_{i,2} \right) = 0.794 \qquad yv := \frac{1}{ny} \cdot \sum_{i=1}^{rows(YN)} \left(YN_{i,1} \cdot YN_{i,2} \right) = 0.703$$

$$Dvx := \frac{1}{nx} \cdot \sum_{i=1}^{rows(XN)} \left[\left(XN_{i,1} - xv \right)^2 \cdot XN_{i,2} \right] = 0.023$$

$$Dvy := \frac{1}{ny} \cdot \sum_{i=1}^{rows(YN)} \left[\left(YN_{i,1} - yv \right)^2 \cdot YN_{i,2} \right] = 0.048$$

$$S2x := \frac{nx}{nx - 1} \cdot Dvx = 0.024 \qquad S2y := \frac{nx}{nx - 1} \cdot Dvy = 0.05$$

$$\alpha := 0.01$$

$$fk := qF(1 - \alpha, ny - 1, nx - 1) = 2.288$$

$$f := \frac{S2y}{S2x} = 2.07$$

<u>Відповідь:</u> при рівні значущості 0,01 дисперсії обох випадкових величин рівні, тобто дисперсії вмісту домішок у олії в обох цехах однакові.

<u>Приклад 5.3.</u>

У відділ стандартизації для дослідження на вміст ГМО поступили дві партії харчової продукції. Для цього було відібрано дві вибірки по 35 одиниць із кожної партії і визначено процентний вміст ГМО. У першій вибірці середній відсоток становив 0,09 % із стандартним відхиленням 0,002 %; у другій – 0,1 % із стандартним відхиленням 0,001 %. Чи є підстави вважати, що процентний вміст ГМО у другій партії більший?

<u>Розв’язання.</u>

Нехай X і Y – процентний вміст ГМО у першій і другій партіях відповідно.

Нульова гіпотеза $H_0: M(X) = M(Y)$ (процентний вміст ГМО у обох партіях однаковий), альтернативна – $H_1: M(X) < M(Y)$ (процентний вміст ГМО у другій партії більший).

Із умови задачі відомо: $n_x = n_y = 35;$ $\bar{x}_в = 0,09;$ $\bar{y}_в = 0,1;$ $\sigma_{вx} = 0,002;$ $\sigma_{вy} = 0,001.$

Оскільки генеральні дисперсії невідомі, то попередньо потрібно встановити, чи рівні між собою генеральні дисперсії $H_0: D_x = D_y$ за альтернативної гіпотези $H_1: D_x \neq D_y$.

Для визначення спостережуваного значення критерію потрібно використати критерій (5.7) $F = \dfrac{S_1^2}{S_2^2}, (S_1^2 > S_2^2),$ де $S_1^2 = S_x^2 = \dfrac{35}{34} \cdot (0,002)^2 \approx$ $\approx 0,000004;$ $S_2^2 = S_y^2 = \dfrac{35}{34} \cdot (0,001)^2 \approx 0,000001:$ $f^* = \dfrac{0,000004}{0,000001} = 4.$

За змістом альтернативної гіпотези потрібно будувати двосторонню критичну область.

Критичне значення $f_{кр}$ знаходиться із таблиці значень розподілу Фішера-Снедекора при рівні значущості $\alpha = 0{,}02$ за умови $f_{кр} = F\left(\frac{\alpha}{2}, k_1, k_2\right)$, де $k_1 = n_1 - 1 = n_y - 1 = 34 = k_2 : k_1 = n_1 - 1 = n_y - 1 = 34 = k_2$:

$$f_{кр} = F\left(\frac{0{,}02}{2}; 34; 34\right) = 1{,}9.$$

Двостороння критична область зображена на рис. 5.3:

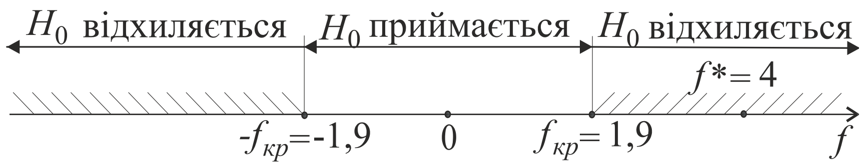


Рис. 5.3

Оскільки $f^* > f_{кр}$ належить критичній області, то гіпотеза H_0 відхиляється, приймається альтернативна гіпотеза H_1: дві генеральні дисперсії різні.

Тепер можна повернутися до перевірки гіпотези про рівність генеральних середніх.

Оскільки генеральні дисперсії невідомі і різні, то для визначення спостережуваного значення критерію потрібно використати критерій (5.6)

$$Z = \frac{\overline{x}_в - \overline{y}_в}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}}: \; z^* = \frac{0{,}09 - 0{,}1}{\sqrt{\frac{0{,}000004}{35} + \frac{0{,}000001}{35}}} \approx -26{,}46.$$

За змістом альтернативної гіпотези потрібно будувати лівосторонню критичну область.

Критичне значення $z_{кр}$ визначається із умови $\Phi(z_{кр}) = -\frac{1 - 2\alpha}{2}$ за таблицею значень функції Лапласа, наприклад, для рівня значущості $\alpha = 0{,}05$:

$$\Phi(z_{кр}) = -\frac{1-2\alpha}{2} = -\frac{1-0,1}{2} = 0,45;\ z_{кр} = -1,645.$$

Лівостороння критична область зображена на рис. 5.4:

Рис. 5.4

Оскільки $z^* < z_{кр}$ належить критичній області, то гіпотеза H_0 відхиляється, приймається альтернативна гіпотеза H_1.

<u>Відповідь:</u> при рівні значущості 0,05 є підстави вважати, що процентний вміст ГМО у другій партії більший.

5.3. Перевірка статистичної гіпотези про рівність часток ознаки у двох сукупностях

На практиці досить часто зустрічається задача про порівняння часток ознаки у двох сукупностях.

Нехай є дві генеральні сукупності, ймовірності появи ознаки (частки ознаки) X у яких дорівнює відповідно p_1 і p_2. Потрібно перевірити нульову гіпотезу про рівність часток ознаки X для цих сукупностей, тобто $H_0: p_1 = p_2$. Для перевірки гіпотези із цих сукупностей взято дві незалежні вибірки обсягами n_1 і n_2. Вибіркові частки появи ознаки дорівнюють відповідно $w_1 = \frac{m_1}{n_1}$ і $w_2 = \frac{m_2}{n_2}$, де m_1, m_2 – число варіант з ознакою X у цих вибірках.

При великих обсягах вибірок розподіл відносних частот наближається до нормального закону розподілу з параметрами: $M(w_1) = p_1$, $M(w_2) = p_2$, $D(w_1) = \dfrac{p_1 q_1}{n_1}$, $D(w_2) = \dfrac{p_2 q_2}{n_2}$, де $q_1 = 1 - p_1$, $q_2 = 1 - p_2$.

За статистичний критерій перевірки нульової гіпотези H_0 береться випадкова величина

$$Z = \frac{w_1 - w_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}, \tag{5.8}$$

Якщо справджується гіпотеза $H_0 : p_1 = p_2 = p,$ то випадкова величина Z має нормальний розподіл із параметрами $M(Z) = 0,$ $D(Z) = 1.$ Дійсно, оскільки $M(w_1 - w_2) = p - p = 0,$ то

$$M(Z) = M\left(\frac{w_1 - w_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right) = \frac{M(w_1) - M(w_2)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 0,$$

$$D(Z) = D\left(\frac{w_1 - w_2}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right) = \frac{D(w_1) + D(w_2)}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{\frac{pq}{n_1} + \frac{pq}{n_2}}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 1.$$

За невідоме значення ймовірності p у формулі (5.8) беруть її найкращу оцінку, що дорівнює вибірковій частці ознаки, коли дві вибірки об'єднати в одну, тобто $p = \dfrac{m_1 + m_2}{n_1 + n_2}$.

Знаходження критичних точок, вибір критичної області проводиться згідно методики, розглянутої раніше.

Приклад 5.4.

На хлібокомбінаті проводиться перевірка системи обліку. Було виявлено, що із 60 відомостей 23 містили неточності. Після усунення недоліків повторна перевірка показала, що із 35 відомостей 5 містили неточності. Чи є підстави вважати, що після усунення помилок частка неточностей зменшилась?

<u>Розв’язання.</u>

Нульова гіпотеза $H_0: p_1 = p_2$ (після усунення помилок частка недоліків не змінилась), альтернативна – $H_1: p_1 > p_2$ (після усунення помилок частка недоліків зменшилась).

Критичне значення $z_{кр}$ визначається із умови $\Phi(z_{кр}) = \dfrac{1-2\alpha}{2}$ за таблицею значень функції Лапласа, наприклад, для рівня значущості $\alpha = 0,01$:

$$\Phi(z_{кр}) = \frac{1-2\alpha}{2} = \frac{1-0,02}{2} = 0,49; \; z_{кр} = 2,98.$$

За змістом альтернативної гіпотези потрібно будувати правосторонню критичну область.

За умовою $n_1 = 60,\ n_2 = 35,\ m_1 = 23,\ m_2 = 5,$ тому

$$p = \frac{m_1 + m_2}{n_1 + n_2} = \frac{23+5}{60+35} \approx 0,29.$$

Для обчислення спостережуваного значення критерію потрібно використати критерій (5.8) $Z = \dfrac{w_1 - w_2}{\sqrt{pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$, де $w_1 = \dfrac{m_1}{n_1} = \dfrac{23}{60} \approx 0,38;$ $w_2 = \dfrac{m_2}{n_2} = \dfrac{5}{35} \approx 0,14;\ q = 1 - p = 1 - 0,29 = 0,71:$

$$z^* = \frac{0,38 - 0,14}{\sqrt{0,29 \cdot 0,71\left(\dfrac{1}{60} + \dfrac{1}{35}\right)}} \approx 2,49.$$

Правостороння критична область зображена на рис. 5.5:

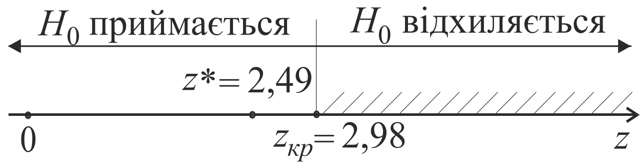


Рис. 5.5

Оскільки $z^* < z_{кр}$ не належить критичній області, то приймається гіпотеза H_0.

<u>Відповідь:</u> при рівні значущості 0,01 є підстави вважати, що після усунення помилок частка недоліків не змінилась.

Критерії перевірки статистичних гіпотез про порівняння параметрів двох сукупностей зведені у табл. 5.1.

Таблиця 5.1

Гіпотеза H_0	**Припущення**	**Статистичний критерій**	**Гіпотеза** H_1	**Критична точка**	**Область при-йняття гіпотези** H_0
$M(X) = M(Y)$	D_x, D_y - відомі, $n_1 > 30, n_2 > 30$	$Z = \dfrac{\overline{x}_в - \overline{y}_в}{\sqrt{\dfrac{D_x}{n_x} + \dfrac{D_y}{n_y}}}$	$M(X) > \mu$ $M(X) < \mu$ $M(X) \neq \mu$	$\Phi(z_{кр}) = \dfrac{1-2\alpha}{2}$ $\Phi(z_{кр}) = -\dfrac{1-2\alpha}{2}$ $\Phi(z_{кр}) = \dfrac{1-\alpha}{2}$	$Z < z_{кр}$ $Z > -z_{кр}$ $\lvert Z\rvert < z_{кр}$
	$D_x \neq D_y$ - невідомі, $n_1 > 30, n_2 > 30$	$Z = \dfrac{\overline{x}_в - \overline{y}_в}{\sqrt{\dfrac{S_x^2}{n_x} + \dfrac{S_y^2}{n_y}}}$			
	$D_x = D_y$ - невідомі, $n_1 > 30, n_2 > 30$	$Z = \dfrac{\overline{x}_в - \overline{y}_в}{S\sqrt{1/n_x + 1/n_y}}$, $S^2 = \dfrac{(n_x - 1)S_x^2 + (n_y - 1)S_y^2}{n_x + n_y - 2}$			
	$D_x = D_y$ - невідомі, $n_1 < 30, n_2 < 30$	$t = \dfrac{\overline{x}_в - \overline{y}_в}{S\sqrt{1/n_1 + 1/n_2}}$, $k = n_1 + n_2 - 2$	$M(X) > \mu$ $M(X) < \mu$ $M(X) \neq \mu$	$t_{кр} = t(\alpha, k)$ $t_{кр} = -t(\alpha, k)$ $t_{кр} = t\left(\dfrac{\alpha}{2}, k\right)$	$t < t_{кр}$ $t > -t_{кр}$ $\lvert t\rvert < t_{кр}$
$\sigma_x^2 = \sigma_y^2$	-	$F = \dfrac{S_1^2}{S_2^2}, (S_1^2 > S_2^2)$	$\sigma_x^2 > \sigma_y^2$ $\sigma_x^2 \neq \sigma_y^2$	$f_{кр} = F(\alpha, k_1, k_2)$ $f_{кр} = F\left(\dfrac{\alpha}{2}, k_1, k_2\right)$	$F < f_{кр}$ $\lvert F\rvert < f_{кр}$
$p_1 = p_2$	$n_1 > 30, n_2 > 30$	$Z = \dfrac{w_1 - w_2}{\sqrt{pq\left(\dfrac{1}{n_1} + \dfrac{1}{n_2}\right)}}$, $p = \dfrac{m_1 + m_2}{n_1 + n_2}$, $w_1 = \dfrac{m_1}{n_1},\ w_2 = \dfrac{m_2}{n_2}$	$p_1 > p_2$ $p_1 < p_2$ $p_1 \neq p_2$	$\Phi(z_{кр}) = \dfrac{1-2\alpha}{2}$ $\Phi(z_{кр}) = -\dfrac{1-2\alpha}{2}$ $\Phi(z_{кр}) = \dfrac{1-\alpha}{2}$	$Z < z_{кр}$ $Z > -z_{кр}$ $\lvert Z\rvert < z_{кр}$

5.4. Перевірка непараметричних статистичних гіпотез. Критерії узгодженості

Однією із важливих задач математичної статистики є встановлення теоретичного закону розподілу ознаки генеральної сукупності. Зокрема, більшість із розглянутих вище критеріїв перевірки параметричних статистичних гіпотез ґрунтувались на припущенні, що ознака генеральної сукупності має нормальний закон розподілу. Отже, використовувати ці критерії потрібно у разі достатньої упевненості, що спостережувана ознака генеральної сукупності має нормальний або близький до нього закон розподілу.

Якщо закон розподілу ознаки невідомий, то на основі дослідження вибіркової сукупності одержують емпіричний розподіл, за яким робиться припущення про вигляд і параметри невідомого розподілу ознаки генеральної сукупності. Припущення про закон розподілу ознаки можна зробити за наявними теоретичними передумовами про характер зміни ознаки. Зокрема, виконання умов центральної граничної теореми свідчить про наявність нормального закону розподілу.

У деяких випадках підставою для висування гіпотези про теоретичний закон розподілу можуть бути деякі формальні властивості статистичного розподілу, наприклад, рівність $\overline{x}_в$ і $\sigma_в$ є ознакою показникового розподілу. Інколи припущення про вигляд теоретичного закону розподілу може ґрунтуватись на графічному зображенні статистичного розподілу вибірки.

Якщо вигляд теоретичного розподілу ознаки встановлений, то за статистичним розподілом вибірки знаходять найкращі оцінки невідомих параметрів цього розподілу.

Потім виникає задача перевірки правильності вибору виду розподілу, узгодженості дійсного теоретичного розподілу з емпіричним. Перевірка гіпотези про узгодженість вибіркового розподілу з теоретичним проводиться на основі спеціально підібраної випадкової величини – *критерію узгодженості*.

Одним із найбільш поширених критеріїв є критерій узгодженості Пірсона χ^2, який визначається за формулою

$$\chi^2 = \sum_{i=1}^{l} \frac{(n_i - m_i)^2}{m_i}, \tag{5.9}$$

де l – число часткових проміжків інтервального статистичного розподілу вибірки, n_i – спостережувана частота для відповідного проміжку, m_i – теоретична частота, знайдена за вибраним розподілом. Для кожного інтервалу $(\alpha_{i-1};\alpha_i)$ теоретичні частоти знаходяться за формулою $m_i = np_i$, де n – обсяг вибірки, $p_i = P(\alpha_{i-1} < X < \alpha_i)$, $i = \overline{1,l}$. Зокрема, для нормального розподілу

$$m_i = n\left[\Phi\left(\frac{\alpha_i - \overline{x}_в}{\sigma_в}\right) - \Phi\left(\frac{\alpha_{i-1} - \overline{x}_в}{\sigma_в}\right)\right]. \tag{5.10}$$

Випадкова величина (5.9) має розподіл χ^2 з $k = l - r - 1$ ступенями свободи (r – число параметрів, якими визначається закон розподілу генеральної сукупності). Наприклад, для розподілу Пуассона, який характеризується одним параметром λ, $r = 1$; для нормального розподілу з двома параметрами μ і σ – $r = 2$.

Якщо усі емпіричні частоти збігаються з теоретичними, то з формули (5.9) випливає, що $\chi^2 = 0$, інакше $\chi^2 > 0$. Отже, критична область завжди буде правосторонньою. Критична точка $\chi^2_{кр}$ знаходиться із таблиці значень розподілу χ^2 при заданому рівні значущості α за умови: $\chi^2_{кр} = \chi^2(\alpha;k)$, $k = l - r - 1$. Нульова гіпотеза H_0 про закон розподілу генеральної сукупності відхиляється, якщо спостережуване значення χ^2_*, визначене за критерієм (5.9) виявиться більшим або рівним $\chi^2_{кр}$, і приймається, якщо $\chi^2_* < \chi^2_{кр}$.

Застосування критерію узгодженості Пірсона χ^2 вимагає дотримання таких умов:

1) обсяг вибірки повинен бути достатньо великий $(n \geq 50)$;

2) інтервали з малочисельними частотами $(n_i < 5)$ попередньо потрібно об'єднати.

Перевірку гіпотези H_0 про те, що генеральна сукупність має нормальний розподіл ознаки X, виконується за схемою:

1) обчислюються $\overline{x}_в, \sigma_в, u_i = \dfrac{\alpha_i - \overline{x}_в}{\sigma_в}, i = \overline{0,l};$

2) обчислюються теоретичні частоти $m_i = n[\Phi(u_i) - \Phi(u_{i-1})]$ за формулою (5.10), $i = \overline{1,l}$, $u_0 = -\infty$, $u_l = +\infty$;

3) обчислюється спостережуване значення критерію χ^2_* за формулою (5.9);

4) визначається критична точка $\chi^2_{кр} = \chi^2(\alpha; k)$, будується правобічна критична область;

5) робиться висновок щодо гіпотези H_0.

Приклад 5.5.

Вимірювалась в'язкість рослинної олії. Результати вимірювання подано у вигляді інтервального статистичного розподілу:

x_i, %	0,5-1	1-1,5	1,5-2	2-2,5	2,5-3	3-3,5
n_i	3	8	16	12	7	4

При рівні значущості $\alpha = 0,05$ перевірити достовірність гіпотези H_0 про нормальний закон розподілу ознаки X – в'язкість рослинної олії.

Розв'язання.

Для обчислення $\overline{x}_в, \sigma_в$ потрібно перейти до дискретного статистичного розподілу:

x_i, %	0,75	1,25	1,75	2,25	2,75	3,25
n_i	3	8	16	12	7	4

$$\overline{x}_в = \frac{1}{50}(0,75\cdot 3 + 1,25\cdot 8 + 1,75\cdot 16 + 2,25\cdot 12 + 2,75\cdot 7 + 3,25\cdot 4) \approx 1,99;$$

$$D_в = \frac{1}{50}((0,75-1,99)^2\cdot 3 + (1,25-1,99)^2\cdot 8 + (1,75-1,99)^2\cdot 16 + (2,25-1,99)^2\cdot 12 +$$

$$+(2,75-1,99)^2\cdot 7 + (3,25-1,99)^2\cdot 4) \approx 0,42;$$

$$\sigma_в = \sqrt{0,42} \approx 0,65.$$

Для перевірки достовірності гіпотези H_0 про нормальний закон розподілу ознаки X – в'язкість рослинної олії використовується критерій узгодженості Пірсона χ^2.

Обчислені теоретичні частоти подано у таблиці 5.2:

Таблиця 5.2

α_{i-1}	α_i	n_i	$u_{i-1} = \frac{\alpha_{i-1} - \overline{x}_в}{\sigma_в}$	$u_i = \frac{\alpha_i - \overline{x}_в}{\sigma_в}$	$\Phi(u_{i-1})$	$\Phi(u_i)$	$m_i = n[\Phi(u_i) - \Phi(u_{i-1})]$
0,5	1	3	-2,29	-1,52	-0,4890	-0,4357	2,7
1	1,5	8	-1,52	-0,75	-0,4357	-0,2734	8,1
1,5	2	16	-0,75	0,02	-0,2734	0,0080	14,1
2	2,5	12	0,02	0,78	0,0080	0,2823	13,7
2,5	3	7	0,78	1,55	0,2823	0,4394	7,9
3	3,5	4	1,55	2,32	0,4394	0,4898	2,5

Розрахунки для обчислення спостережуваного значення критерію приведено у таблиці 5.3:

Таблиця 5.3

$n_i - m_i$	$(n_i - m_i)^2$	$\frac{(n_i - m_i)^2}{m_i}$
0,3	0,09	0,03
-0,1	0,01	0,00
1,9	3,61	0,26
-1,7	2,89	0,21
-0,9	0,81	0,10
1,5	2,25	0,90

Тоді $\chi_*^2 = \sum_{i=1}^{6} \frac{(n_i - m_i)^2}{m_i} \approx 1,5.$

За таблицею (додаток 7) при відомих $\alpha = 0,05$, $k = 6 - 2 - 1 = 3$ для розподілу χ^2: $\chi_{кр}^2 = \chi^2(\alpha; k) = \chi^2(0,05; 3) = 7,8.$

Оскільки $\chi_*^2 < \chi_{кр}^2$ (рис. 5.6), то гіпотеза H_0 приймається: є підстави вважати, що ознака X (в'язкість рослинної олії) має нормальний розподіл.

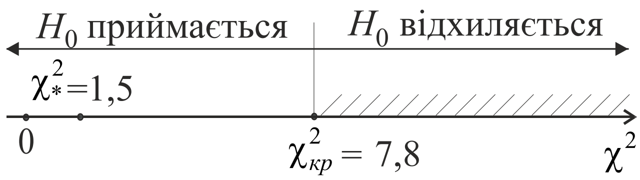


Рис. 5.6

$$ORIGIN := 1$$

$$IN := \begin{pmatrix} 0.5 & 1 & 3 \\ 1 & 1.5 & 8 \\ 1.5 & 2 & 16 \\ 2 & 2.5 & 12 \\ 2.5 & 3 & 7 \\ 3 & 3.5 & 4 \end{pmatrix} \qquad n := \sum IN^{\langle 3 \rangle} = 50$$

$$X := \left| \begin{array}{l} \text{for } i \in 1..\text{rows}(IN) \\ \quad x_{i,1} \leftarrow \dfrac{IN_{i,1} + IN_{i,2}}{2} \\ x \end{array} \right. \qquad X = \begin{pmatrix} 0.75 \\ 1.25 \\ 1.75 \\ 2.25 \\ 2.75 \\ 3.25 \end{pmatrix}$$

$$xv := \frac{1}{n} \cdot \sum_{i=1}^{\text{rows}(X)} \left(X_{i,1} \cdot IN_{i,3} \right) = 1.99$$

$$Dv := \frac{1}{n} \cdot \sum_{i=1}^{\text{rows}(X)} \left[\left(X_{i,1} - xv \right)^2 \cdot IN_{i,3} \right] = 0.422$$

$$\sigma v := \sqrt{Dv} = 0.65$$

$$u1 := \left| \begin{array}{l} \text{for } i \in 1..\text{rows}(IN) \\ \quad u1_{i,1} \leftarrow \dfrac{IN_{i,1} - xv}{\sigma v} \\ u1 \end{array} \right. \qquad u1 = \begin{pmatrix} -2.293 \\ -1.523 \\ -0.754 \\ 0.015 \\ 0.785 \\ 1.554 \end{pmatrix}$$

$$u2 := \left| \begin{array}{l} \text{for } i \in 1..\text{rows}(IN) \\ \quad u2_{i,1} \leftarrow \dfrac{IN_{i,2} - xv}{\sigma v} \\ u2 \end{array} \right. \qquad u2 = \begin{pmatrix} -1.523 \\ -0.754 \\ 0.015 \\ 0.785 \\ 1.554 \\ 2.323 \end{pmatrix}$$

$$\Phi 1 := \text{pnorm}(u1, 0, 1) - 0.5$$

$$\Phi 2 := \text{pnorm}(u2, 0, 1) - 0.5$$

$$\Phi 2^T = \begin{pmatrix} -0.436 & -0.275 & 6.138 \times 10^{-3} & 0.284 & 0.44 & 0.49 \end{pmatrix}$$

$$M := n \cdot (\Phi 2 - \Phi 1) \qquad M^T = \begin{pmatrix} 2.646 & 8.08 & 14.035 & 13.877 & 7.811 & 2.5 \end{pmatrix}$$

$$\chi 2 := \sum_{i=1}^{\text{rows}(IN)} \frac{\left(IN_{i,3} - M_{i,1}\right)^2}{M_{i,1}} = 1.561$$

$$\alpha := 0.05$$

$$\chi 2k := \text{qchisq}(1 - \alpha, 3) = 7.815$$

Відповідь: при рівні значущості $\alpha = 0{,}05$ є підстави вважати, що ознака X – в'язкість рослинної олії має нормальний розподіл.

Приклад 5.6.

Задано статистичний розподіл вибірки:

$[\alpha_{i-1};\alpha_i)$	0-10	10-20	20-30	30-40	40-50
n_i	50	35	15	7	3

Визначити теоретичний закон розподілу ознаки генеральної сукупності X та при рівні значущості $\alpha = 0,01$ перевірити правильність висунутої нульової гіпотези.

Розв’язання.

Гістограму частот заданого статистичного розподілу зображено на рис. 5.7, із умови $h = 10$; $f_1 = 5$; $f_2 = 3,5$; $f_3 = 1,5$; $f_4 = 0,7$; $f_5 = 0,3$:

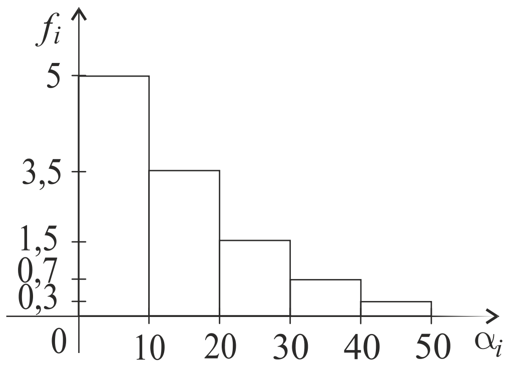


Рис. 5.7

За формою отриманої гістограми можна припустити, що ознака X має показниковий розподіл. Для перевірки припущення можна застосувати критерій Пірсона. Теоретичні частоти обчислюються за формулою:

$$m_i = n(e^{-\lambda\alpha_{i-1}} - e^{-\lambda\alpha_i}),$$

де $\lambda = \dfrac{1}{\overline{x}_в}$.

Для обчислення $\overline{x}_в$ потрібно перейти до дискретного статистичного розподілу:

x_i, %	5	15	25	35	45
	50	35	15	7	3

$\overline{x}_в = \dfrac{1}{110}(5\cdot 50 + 15\cdot 35 + 25\cdot 15 + 35\cdot 7 + 45\cdot 3) \approx 13,91;$ тоді $\lambda = \dfrac{1}{\overline{x}_в} = \dfrac{1}{13,91} \approx 0,072.$

Обчислені теоретичні частоти подано у таблиці 5.4:

Таблиця 5.4

α_{i-1}	α_i	n_i	$e^{-\lambda\alpha_{i-1}}$	$e^{-\lambda\alpha_i}$	$m_i = n(e^{-\lambda\alpha_{i-1}} - e^{-\lambda\alpha_i})$
0	10	50	1	0,4873	56
10	20	35	0,4873	0,2374	27
20	30	15	0,2374	0,1157	13
30	40	7	0,1157	0,0564	7
40	50	3	0,0564	0,0275	3

Розрахунки для обчислення спостережуваного значення критерію приведено у таблиці 5.5:

Таблиця 5.5

$n_i - m_i$	$(n_i - m_i)^2$	$\frac{(n_i - m_i)^2}{m_i}$
-6	36	0,6429
8	64	2,3704
2	4	0,3077
0	0	0
0	0	0

Тоді $\chi_*^2 = \sum_{i=1}^{5} \frac{(n_i - m_i)^2}{m_i} \approx 3.$

За таблицею (додаток 7) при відомих $\alpha = 0,01$, $k = 5-1-1 = 3$ для розподілу χ^2: $\chi_{кр}^2 = \chi^2(\alpha; k) = \chi^2(0,01; 3) = 11,3.$

Оскільки $\chi_*^2 < \chi_{кр}^2$ (рис. 5.8), то гіпотеза H_0 приймається, є підстави вважати, що ознака X має показниковий розподіл.

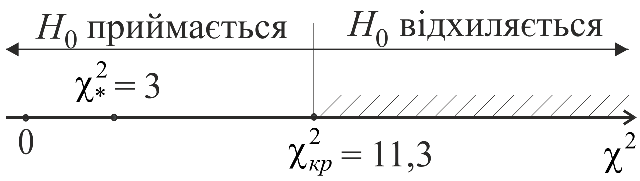


Рис. 5.8

$\text{ORIGIN} := 1$

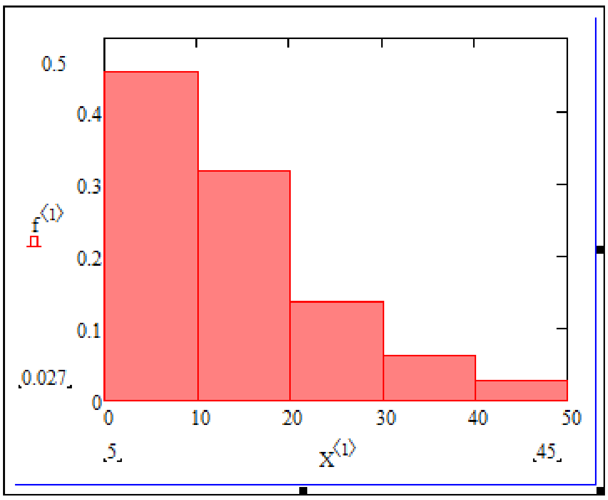
$$\text{IN} := \begin{pmatrix} 0 & 10 & 50 \\ 10 & 20 & 35 \\ 20 & 30 & 15 \\ 30 & 40 & 7 \\ 40 & 50 & 3 \end{pmatrix}$$

$$n := \sum \text{IN}^{\langle 3 \rangle} = 110$$

$$X := \left| \begin{array}{l} \text{for } i \in 1..\text{rows}(\text{IN}) \\ \quad x_{i,1} \leftarrow \dfrac{\text{IN}_{i,1} + \text{IN}_{i,2}}{2} \\ x \end{array} \right. \qquad X = \begin{pmatrix} 5 \\ 15 \\ 25 \\ 35 \\ 45 \end{pmatrix}$$

$$f := \left| \begin{array}{l} \text{for } i \in 1..\text{rows}(\text{IN}) \\ \quad g_{i,1} \leftarrow \dfrac{\text{IN}_{i,3}}{n} \\ g \end{array} \right. \qquad f = \begin{pmatrix} 0.455 \\ 0.318 \\ 0.136 \\ 0.064 \\ 0.027 \end{pmatrix}$$

$$X := \left| \begin{array}{l} \text{for } i \in 1..\text{rows}(\text{IN}) \\ \quad x_{i,1} \leftarrow \dfrac{\text{IN}_{i,1} + \text{IN}_{i,2}}{2} \\ x \end{array} \right. \qquad X = \begin{pmatrix} 5 \\ 15 \\ 25 \\ 35 \\ 45 \end{pmatrix}$$



$$xv := \frac{1}{n} \cdot \sum_{i=1}^{\mathrm{rows}(IN)} \left(X_{i,1} \cdot IN_{i,3}\right) = 13.909$$

$$\lambda := \frac{1}{xv} = 0.072$$

$$\alpha 1 := IN^{\langle 1 \rangle} \qquad \alpha 2 := IN^{\langle 2 \rangle}$$

$$P1 := e^{-\lambda \cdot \alpha 1} \qquad P2 := e^{-\lambda \cdot \alpha 2}$$

$$P1^{T} = (1 \;\; 0.487 \;\; 0.237 \;\; 0.116 \;\; 0.056)$$

$$P2^{T} = (0.487 \;\; 0.237 \;\; 0.116 \;\; 0.056 \;\; 0.027)$$

$$M := n \cdot (P1 - P2)$$

$$\chi 2 := \sum_{i=1}^{\mathrm{rows}(IN)} \frac{\left(IN_{i,3} - M_{i,1}\right)^2}{M_{i,1}} = 3.021$$

$$\alpha := 0.01$$

$$\chi 2k := \mathrm{qchisq}(1 - \alpha, 3) = 11.345$$

Відповідь: при рівні значущості $\alpha = 0,01$ є підстави вважати, що ознака X має показниковий розподіл.

Практичне заняття 5

5.1. Два бухгалтери працюють в одному відділі. Для аналізу часу, який кожен з них витрачає на заповнення рахунку-фактури було проведене дослідження і отримані результати: перший бухгалтер – $n_x = 35$, $\bar{x}_в = 8,5$ (хв.), $D_x = 1$; другий – $n_y = 31$, $\bar{y}_в = 9$ (хв.), $D_y = 0,5$. Чи можна за рівня значущості

$\alpha = 0,01$ стверджувати, що перший бухгалтер заповнює рахунки-фактури швидше, ніж другий?

5.2. Менеджер мережі продуктових магазинів домовився із двома постачальниками круп. Вирішивши перевірити обох постачальників на відповідність круп за вагою, було проведене дослідження і отримані результати: перший постачальник – $n_x = 40$, $\overline{x}_в = 1000$ (г), $D_x = 1$; другий – $n_y = 40$, $\overline{y}_в = 1001$ (г), $D_y = 2$. Чи є підстави вважати суттєвою розбіжність між вагою круп двох постачальників (рівень значущості $\alpha = 0,05$)?

5.3. У відділ стандартизації поступили дві партії харчової продукції для дослідження на вміст шкідливих домішок. Для цього було відібрано дві вибірки по 40 одиниць із кожної партії і визначено процентний вміст шкідливих домішок. У першій вибірці середній відсоток становив 0,1 % із стандартним відхиленням 0,001 %; у другій – 0,15 % із стандартним відхиленням 0,01 %. Чи є підстави вважати, що процентний вміст домішок у першій партії більший?

5.4. Керівництво концерну вирішило проаналізувати прибутки двох підприємств одного регіону за певний період. У результаті було отримано вибірки:

x_i, тис. ум. од.	2,2	2,4	2,6	2,8
n_{x_i}	4	15	10	3

y_i, тис. ум. од.	2,1	2,4	2,7	3
n_{y_i}	5	17	8	1

Чи можна за рівня значущості $\alpha = 0,0$ вважати, що прибутки першого підприємства менші, ніж другого?

5.5. Для контролю температурного режиму у процесі виготовлення йогуртів було проведено заміри температури: 36; 35,5; 36; 35; 36; 35; 36,5. З метою регулювання температурного режиму відбулося налагодження термостату і було проведено нові заміри температури: 35; 37; 35,5; 37; 37; 36. Чи є підстави вважати суттєвою розбіжність між температурними режимами до і після налагодження термостату (рівень значущості $\alpha = 0,05$)?

5.6. На підприємство по виробництву олії поступило дві партії насіння соняшника, які згодом були відправлені на окремі технологічні лінії. Відділом контролю була проведена перевірка на вміст домішок у отриманій олії. Результати дослідження подано у вигляді двох розподілів:

x_i, %	0,3	0,5	0,7	0,9
n_{x_i}	3	15	11	1

y_j, %	0,1	0,4	0,7	1,0
n_{y_j}	2	17	15	1

При рівні значущості $\alpha = 0,01$ встановити чи є друга партія насіння більш якісна, ніж перша?

5.7. На підприємстві працюють дві лінії по фасуванню майонезу. Відділом контролю була проведена перевірка ваги пачок, фасованих по 200 г на першій та другій лініях. Результати перевірки подано у вигляді двох розподілів:

x_i, г	198	199	200	201
n_{x_i}	3	15	17	2

y_j, г	197	199	200	201
n_{y_j}	2	16	18	3

Ознаки X, Y (вага пачок майонезу) є незалежними нормально розподіленими випадковими величинами. При рівні значущості $\alpha = 0,05$ перевірити гіпотезу $H_0 : D_x = D_y$ за альтернативної $H_1 : D_x > D_y$.

5.8. Відомо, що I інвестиція розрахована на 11 років, дисперсія щорічних прибутків складає 12 %, II інвестиція розрахована на 12 років, дисперсія щорічних прибутків складає 17 %. Припускається, що розподіл щорічних прибутків на інвестиції є нормальним. Дослідити чи рівні ризики при рівні значущості $\alpha = 0,01$?

5.9. Вивчались показники діяльності підприємств харчової промисловості деяких двох регіонів. Дані про обсяг реалізованої продукції подані у вигляді статистичних розподілів:

Інтервал, млн.ум.од.	0,5-2	2-3,5	3,5-5	5-6,5	6,5-8
n_{x_i}	2	20	22	15	3

Інтервал, млн.ум.од.	1-3,5	3,5-6	6,5-9	9-11,5
n_{y_j}	5	25	15	7

При рівні значущості $\alpha = 0,05$ перевірити гіпотезу $H_0 : D_x = D_y$ за альтернативної $H_1 : D_x \neq D_y$.

5.10. Аудитор проводить перевірку на підприємстві. У результаті виявилось, що з 200 навмання відібраних для перевірки звітів 30 містили неточності. Працівникам економічного відділу підприємства було запропоновано виправити недоліки, після чого знову була здійснена перевірка. Із 230 навмання відібраних для перевірки звітів 35 містили неточності. Чи є підстави вважати, що працівники підприємства покращили звітність? Рівень значущості $\alpha = 0,01$.

5.11. Для перевірки засвоєння матеріалу двох розділів дисципліни «Управління персоналом» студенту було запропоновано пройти тести. Із 150 питань тесту, складеного за матеріалами першого розділу, студент відповів правильно на 120 питань, із 120 питань тесту, складеного за другим розділом, – на 90. 1) Чи є підстави вважати суттєвою розбіжність у засвоюванні обох розділів студентом? 2) Чи є підстави стверджувати, що перший розділ студент засвоїв гірше, ніж другий? Рівень значущості $\alpha = 0,05$.

5.12. На підприємство було закуплено два види енергозберігаючих лампочок. Через деякий час з’ясувалось, що із 80 лампочок першого виду 5 перестали світити, а із 60 другого – 2 перестали світити. Якого виду лампочки краще придбати наступного разу? Рівень значущості $\alpha = 0,01$.

5.13. Задано статистичний розподіл вибірки:

1)

$[\alpha_{i-1};\alpha_i)$	50-60	60-70	70-80	80-90	90-100
n_i	1	12	50	20	5

2)

$[\alpha_{i-1};\alpha_i)$	0,5-6,5	6,5-12,5	12,5-18,5	18,5-24,5	24,5-30,5	30,5-36,5
n_i	7	11	30	35	7	4

Визначити теоретичний закон розподілу ознаки генеральної сукупності X та при рівні значущості $\alpha = 0,01$ перевірити правильність висунутої нульової гіпотези.

5.14. Вивчались показники інноваційної діяльності 50 підприємств харчової промисловості деякого регіону. Дані про обсяг реалізованої інноваційної продукції подані у вигляді статистичного розподілу:

Інтервал, тис.ум.од.	1-3	3-5	5-7	7-9	9-11
n_i	3	15	53	22	7

При рівні значущості $\alpha = 0,05$ перевірити достовірність гіпотези H_0 про нормальний закон розподілу ознаки X – обсяг реалізованої інноваційної продукції.

5.15. Задано статистичний розподіл вибірки:

1)

$[\alpha_{i-1};\alpha_i)$	10-20	20-30	30-40	40-50	50-60
n_i	50	35	15	7	3

2)

$[\alpha_{i-1};\alpha_i)$	0-6	6-12	12-18	18-24	24-30
n_i	40	30	20	5	2

Визначити теоретичний закон розподілу ознаки генеральної сукупності X та при рівні значущості $\alpha = 0,01$ перевірити правильність висунутої нульової гіпотези.

5.16. На підприємство на переробку надходять огірки. Відділом контролю було перевірено вміст нітратів (ум. од.) в огірках. Результати наведено у таблиці:

$[\alpha_{i-1};\alpha_i)$	2-4,5	4,5-7,0	7,0-9,5	9,5-12,0	12,0-14,5
n_i	50	35	15	7	3

При рівні значущості $\alpha = 0,05$ перевірити достовірність гіпотези H_0 про показниковий закон розподілу ознаки X – вміст нітратів.

Лекція 6. Основи кореляційного та регресійного аналізу

6.1. Двовимірний статистичний розподіл вибірки та його числові характеристики

6.2. Поняття статистичної та кореляційної залежностей

6.3. Лінійна парна регресія

6.4. Коефіцієнт кореляції вибірки та його властивості

Кореляційний і регресійний аналіз широко використовуються у різних областях науки і практики при дослідженні взаємозв’язку між кількома величинами або явищами за наявними статистичними даними. Поняття кореляції (від латинського «correlatio» – співвідношення, взаємозв’язок) і регресії (від латинського «regressio» – рух назад) введені у XIX ст. англійськими статистами Ф. Гальтоном і К. Пірсоном.

6.1. Двовимірний статистичний розподіл вибірки та його числові характеристики

У багатьох задачах різних областей знань, зокрема економіки, часто доводиться вивчати взаємозв’язок між двома випадковими величинами. Наприклад, розглянемо підприємство, яке регулярно проводить курси підвищення кваліфікації для свого персоналу. Щомісячно підприємством фіксуються суми коштів, витрачених на підвищення кваліфікації персоналу та одержаних від реалізації продукції. Є підстави вважати, що повинен існувати якийсь зв’язок між витратами на підвищення кваліфікації персоналу (X) і відповідними щомісячними обсягами реалізації (Y). Можна вважати, що зростання витрат на підвищення кваліфікації, має привести до зростання обсягу реалізації. Закону ж, який би давав можливість записати рівняння, що відображає залежність між X та Y, не існує. Взагалі є ряд факторів, пов’язаних між собою (ціна продукції, ціна конкурентної продукції, чисельність населення, період часу тощо), які точно визначають щомісячний обсяг реалізації. Якщо витрати на підвищення кваліфікації персоналу впливають на обсяг реалізації, то знання про наявність зв’язку між цими двома змінними і його вигляд є досить важливим для оцінки обсягу реалізації і планування діяльності підприємства.

Нехай в генеральній сукупності досліджуються кількісні ознаки двох випадкових величин X та Y. З цією метою відібрано вибірку обсягом n, в якій спостерігалися значення $x_1, x_2, ..., x_k$ ознаки X і значення $y_1, y_2, ..., y_m$ ознаки Y. Кожне значення $X = x_i$ може спостерігатися із будь-яким значенням $Y = y_i$ і частота появи пари $(x_i; y_j)$ становить n_{ij}, $i = \overline{1,k}$, $j = \overline{1,m}$. Одержаний двовимірний статистичний розподіл вибірки представимо у вигляді *кореляційної* таблиці (табл. 6.1):

Таблиця 6.1

y_j \ x_i	y_1	…	y_j	…	y_m	n_x
x_1	n_{11}	…	n_{1j}	…	n_{1m}	n_{x_1}
…	…	…	…	…	…	…
x_i	n_{i1}	…	n_{ij}	…	n_{im}	n_{x_j}
…	…	…	…	…	…	…
x_k	n_{k1}	…	n_{kj}	…	n_{km}	n_{x_k}
n_y	n_{y_1}	…	n_{y_j}	…	n_{y_m}	n

В останньому рядку та стовпці таблиці обчислені суми частот у відповідних рядках та стовпцях:

$$n_{x_i} = \sum_{j=1}^{m} n_{ij},\, n_{y_j} = \sum_{i=1}^{k} n_{ij},\, i = \overline{1,k},\ j = \overline{1,m},$$

при цьому

$$\sum_{i=1}^{k} n_{x_i} = \sum_{j=1}^{m} n_{y_j} = \sum_{i=1}^{k}\sum_{j=1}^{m} n_{ij} = n \text{ – обсяг вибірки.}$$

Загальні середні $\overline{x}$, $\overline{y}$ відповідно змінних X та Y визначаються за формулами:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{m} x_i n_{ij} = \frac{1}{n}\sum_{i=1}^{k} x_i n_{x_i}, \qquad (6.1)$$

$$\overline{y} = \frac{1}{n}\sum_{i=1}^{k}\sum_{j=1}^{m} y_j n_{ij} = \frac{1}{n}\sum_{j=1}^{m} y_j n_{y_j}. \qquad (6.2)$$

Дисперсія $\widehat{D}_x$ змінної X відносно її загальної середньої $\overline{x}$ обчислюється за формулою

$$\widehat{D}_x = \widehat{\sigma}_x^2 = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x})^2 n_{x_i} = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_{x_i} - (\overline{x})^2. \qquad (6.3)$$

Аналогічно дисперсія змінної Y відносно її загальної середньої $\overline{y}$ є

$$\widehat{D}_y = \widehat{\sigma}_y^2 = \frac{1}{n}\sum_{j=1}^{m}(y_i - \overline{y})^2 n_{y_j} = \frac{1}{n}\sum_{j=1}^{m} y_i^2 n_{y_j} - (\overline{y})^2. \qquad (6.4)$$

Двовимірний статистичний розподіл вибірки можна розбити на групи. Розглянемо розподіл ознаки Y при фіксованому значенні $X = x_i$. Такий статистичний розподіл називається умовним. Отже, вся сукупність розбивається на k груп, кожна з яких відрізняється фіксованим значенням x_i, $i = \overline{1,k}$. Умовний розподіл ознаки Y при фіксованому значенні $X = x_i$ має вигляд (табл. 6.2):

Таблиця 6.2

Y	y_1	…	y_j	…	y_m
n_{ij}	n_{i1}	…	n_{ij}	…	n_{im}

Числові характеристики такого розподілу ознаки Y називаються *умовними* і позначаються $\overline{y}_x$ – умовна середня, $\widehat{D}_{y/x_i}$ – умовна дисперсія, $\widehat{\sigma}_{y/x_i}$ – умовне середнє квадратичне відхилення. Вони обчислюються за формулами:

$$\overline{y}_{x_i} = \frac{1}{n_{x_i}} \sum_{j=1}^{m} y_j n_{ij}, \tag{6.5}$$

$$\widehat{D}_{y/x_i} = \frac{1}{n_{x_i}} \sum_{j=1}^{m} y_j^2 n_{ij} - (\overline{y}_{x_i})^2, \tag{6.6}$$

$$\widehat{\sigma}_{y/x_i} = \sqrt{\widehat{D}_{y/x_i}}. \tag{6.7}$$

Аналогічно сукупність можна розбити на m груп, в кожній з яких фіксується значення $Y = y_j$, $j = \overline{1,m}$. Отриманий умовний статистичний розподіл ознаки X при фіксованому значенні $Y = y_j$ можна теж подати у вигляді таблиці (табл. 6.3):

Таблиця 6.3

X	x_1	...	x_j	...	x_k
n_{ij}	n_{1j}	...	n_{ij}	...	n_{kj}

Умовні числові характеристики ознаки X обчислюються аналогічно, як і для ознаки Y:

$$\overline{x}_{y_j} = \frac{1}{n_{y_j}} \sum_{i=1}^{k} x_i n_{ij}, \tag{6.8}$$

$$\widehat{D}_{x/y_j} = \frac{1}{n_{y_j}} \sum_{i=1}^{k} x_i^2 n_{ij} - (\overline{x}_{y_j})^2, \tag{6.9}$$

$$\widehat{\sigma}_{x/y_j} = \sqrt{\widehat{D}_{x/y_j}}. \tag{6.10}$$

Якщо відомі значення умовних середніх $\overline{y}_{x_i}$, $\overline{x}_{y_j}$ усіх груп, то загальні середні ознаки X та Y можна обчислити за формулами:

$$\overline{x} = \frac{1}{n} \sum_{j=1}^{m} \overline{x}_{y_j} n_{y_j}, \tag{6.11}$$

$$\overline{y} = \frac{1}{n} \sum_{i=1}^{k} \overline{y}_{x_i} n_{x_i}. \tag{6.12}$$

Зауваження. Якщо ознаки X та Y мають інтервальний двовимірний статистичний розподіл, то в кореляційній таблиці за x_i та y_j приймаються середини відповідних інтервалів.

6.2. Поняття статистичної та кореляційної залежностей

До цих пір розглядалася *функціональна* залежність $y = \varphi(x)$ між змінними X та Y, коли кожному значенню $X = x$ ставилося у відповідність єдине значення $Y = y$. Така залежність в певній мірі є абстракцією, а в дійсності величини X та Y знаходяться під впливом багатьох факторів. Це призводить до того, що кожному значенню однієї змінної відповідає деяка множина можливих значень іншої змінної. Інакше кажучи, кожному значенню $X = x$ відповідає деякий умовний розподіл величини Y. Така залежність називається *стохастичною* або *статистичною*. Її частинним випадком є кореляційна залежність.

Означення. Кореляційною залежністю між двома величинами називається функціональна залежність між значеннями однієї із них і умовним середнім другої. Кореляційна залежність двох змінних називається *парною* і може бути представлена у вигляді:

$$\bar{y}_x = f(x), \tag{6.13}$$

або

$$\bar{x}_y = g(y). \tag{6.14}$$

Рівняння (6.13), (6.14) називаються *рівняннями регресії* відповідно Y на X і X на Y, функції $f(x)$, $g(y)$ – *функціями регресії*, а їхні графіки – *лініями регресії*.

При вивченні статистичної залежності між випадковими величинами розглядаються дві основні задачі:

1) *задача регресійного аналізу* – встановлення вигляду рівняння регресії (6.13) або (6.14) (лінійне, квадратичне, показникові тощо);

2) *задача кореляційного аналізу* – оцінка сили або тісноти кореляційної залежності між випадковими величинами X та Y через величину розсіювання значень X або Y навколо лінії регресії. Велике розсіювання свідчить про слабкий кореляційний зв'язок і навпаки.

6.3. Лінійна парна регресія

Нехай із генеральної сукупності зроблено вибірку обсягом n, результати представлені у вигляді кореляційної табл. 6.1.

Графічне зображення статистичної залежності між ознаками X та Y у вигляді точок на координатній площині Oxy називається *полем кореляції* (рис. 6.1).

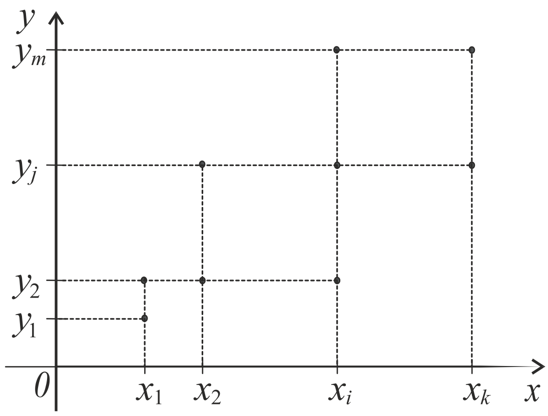


Рис. 6.1

Для кожного умовного розподілу випадкової величини Y при фіксованому значенні величини $X = x_i$ (табл. 6.1) за формулою (6.5) обчислюються умовні середні $\bar{y}_{x_i}, i = \overline{1,k}$. Знайдені значення представлені у вигляді табл. 6.4.

Таблиця 6.4

x_i	x_1	…	x_k
$\overline{y}_{x_i}$	$\overline{y}_{x_1}$	…	$\overline{y}_{x_k}$

На координатній площині наносяться точки $M_i(x_i, \overline{y}_{x_i})$ (рис. 6.2).

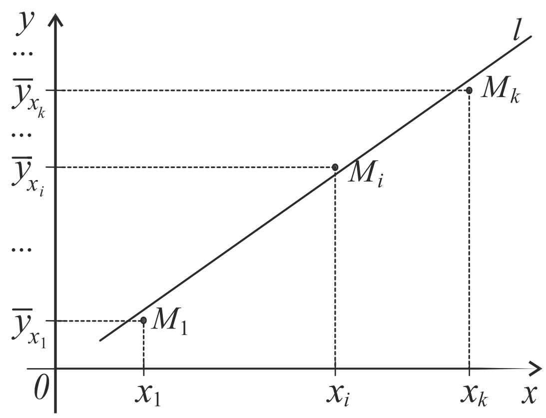


Рис. 6.2

Нехай на підставі візуального аналізу встановлено, що між змінними x та y існує лінійна кореляційна залежність Y на X. Якщо всі точки $M_i, i = \overline{1,k}$ лежать на деякій прямій l, то це буде шукана пряма, рівняння якої знаходиться за двома точками. Але на справді окремі значення $\overline{y}_{x_i}$ можуть відхилятися від прямої l внаслідок дії випадкових факторів. Враховуючи можливі відхилення, парну лінійну регресію можна представити у вигляді:

$$\overline{y}_x = \alpha x + \beta + \varepsilon, \tag{6.15}$$

де α, β – невідомі параметри (коефіцієнти) регресії, ε – випадкове збурення, яке характеризує відхилення від обраної (теоретичної) регресії. Вважається, що значення випадкової величини ε є нормально розподіленою з параметрами $N(0;\sigma_\varepsilon)$ і послідовність значень ε не залежать один від одного.

Отже, права частина рівняння (6.15) є сумою двох частин: систематичної (детермінованої) $\tilde{y} = \alpha x + \beta$ і випадкової ε. Невідомі параметри α, β визначаються за вибіркою, але вони не істинні, оскільки вибірка має обмежений обсяг. Нехай знайдені значення a, b параметрів, вони будуть статистичними оцінками істинних параметрів регресії, тобто рівняння парної регресії

$$y^* = ax + b \tag{6.16}$$

є оцінкою моделі (6.15).

Статистичні оцінки a, b параметрів регресії α, β знаходяться за методом найменших квадратів. При $x = x_i$, рівняння (6.15) набуде вигляду

$$\bar{y}_{x_i} = \alpha x_i + \beta + \varepsilon_i,\ i = \overline{1,k}.$$

Звідси, сума квадратів відхилень від лінії регресії l буде функцією параметрів α, β:

$$Q = \sum_{i=1}^{k} \varepsilon_i^2 n_{x_i} = \sum_{i=1}^{k} (\bar{y}_{x_i} - \alpha x_i - \beta)^2 n_{x_i}. \tag{6.17}$$

Згідно метода найменших квадратів, потрібно так підібрати оцінки параметрів α та β, щоб значення функції $Q(\alpha,\beta)$ було мінімальним.

Необхідною умовою існування екстремуму функції (6.17) є рівність нулю частинних похідних першого порядку по параметрах:

$$\begin{cases} \left.\dfrac{\partial Q}{\partial \alpha}\right|_{\alpha=a} = -2\sum\limits_{i=1}^{k} x_i (\bar{y}_{x_i} - a x_i - b) n_{x_i} = 0; \\ \left.\dfrac{\partial Q}{\partial \beta}\right|_{\beta=b} = -2\sum\limits_{i=1}^{k} (\bar{y}_{x_i} - a x_i - b) n_{x_i} = 0. \end{cases}$$

Розкривши дужки і врахувавши формули (6.1), (6.2), можна прийти до системи лінійних алгебраїчних рівнянь відносно оцінок a та b:

$$\begin{cases} a\sum_{i=1}^{k} x_i^2 n_{x_i} + b\sum_{i=1}^{k} x_i n_{x_i} = \sum_{i=1}^{k} \overline{y}_{x_i} x_i n_{x_i}; \\ a\sum_{i=1}^{k} x_i n_{x_i} + b\sum_{i=1}^{k} n_{x_i} = \sum_{i=1}^{k} \overline{y}_{x_i} n_{x_i}; \end{cases}$$

або

$$\begin{cases} a\overline{x}^2 + b\overline{x} = \overline{xy}; \\ a\overline{x} + b = \overline{y}; \end{cases} \qquad (6.18)$$

де середні $\overline{x}^2$, $\overline{xy}$ обчислюються за формулами

$$\overline{x}^2 = \frac{1}{n}\sum_{i=1}^{k} x_i^2 n_{x_i},\ \overline{xy} = \sum_{i=1}^{k} \overline{y}_{x_i} x_i n_{x_i}. \qquad (6.19)$$

Із системи (6.18) можна знайти статистичні оцінки a, b:

$$\begin{cases} b = \overline{y} - a\overline{x}; \\ a = \dfrac{\overline{xy} - \overline{x}\cdot\overline{y}}{\overline{x}^2 - (\overline{x})^2}. \end{cases} \qquad (6.20)$$

Коефіцієнт a називається *коефіцієнтом регресії* Y на X і позначається символом ρ_{yx}. Його можна записати у вигляді:

$$\rho_{xy} = a = \frac{\overline{xy} - \overline{x}\cdot\overline{y}}{\widehat{\sigma}_x^2}, \qquad (6.21)$$

де $\widehat{\sigma}_x^2 = \overline{x}^2 - (\overline{x})^2$ – дисперсія змінної X вибірки.

Підставивши значення b із системи (6.20) у рівняння (6.16) і врахувавши (6.21), а також всі розрахунки, можна записати рівняння регресії Y на X у вигляді:

$$y^* = \rho_{yx}(x - \overline{x}) + \overline{y}. \qquad (6.22)$$

Аналогічно, вважаючи рівняння регресії (6.14) лінійним, можна привести його до вигляду

$$x^* = \rho_{xy}(y - \overline{y}) + \overline{x}, \qquad (6.23)$$

де ρ_{xy} – коефіцієнт регресії X на Y:

$$\rho_{xy} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{y^2} - (\overline{y})^2} = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\widehat{\sigma}_y^2}. \qquad (6.24)$$

<u>*Зауваження.*</u> Якщо при дослідженні кореляційної залежності між змінними X та Y одержано вибірку $(x_i; y_i)$, обсягу n та $n_{ij} = 0$ при $i \neq j$ і $n_{ij} = 1$ при $i = \overline{1,n}$, $j = \overline{1,n}$, то кореляційна таблиця матиме вигляд (табл. 6.5):

Таблиця 6.5

X	x_1	x_2	$\dots$	x_n
Y	y_1	y_2	$\dots$	y_n

У цьому випадку у відповідних формулах потрібно приймати $n_{ij} = n_{x_i} = n_{y_j} = 1$, а суми $\sum_{i=1}^{k}$, $\sum_{j=1}^{m}$ замінити на $\sum_{i=1}^{n}$.

6.4. Коефіцієнт кореляції вибірки та його властивості

Підбір моделі на основі геометричного аналізу не завжди є вірним, оскільки розміщення точок на графіку залежить від вибраного масштабу. Після вибору функції як форми кореляційної залежності між змінними X та Y потрібно розв'язати другу проблему кореляційного аналізу: оцінити *силу* або *тісноту* такого зв'язку, виходячи із оцінки розсіювання відносно лінії регресії однієї змінної для різних значень другої.

Показником корельованості (лінійного зв'язку) випадкових величин X та Y є *коефіцієнт кореляції* r вибірки, який обчислюється за формулою

$$r = \frac{\overline{xy} - \overline{x} \cdot \overline{y}}{\widehat{\sigma}_x \cdot \widehat{\sigma}_y}. \qquad (6.25)$$

Коефіцієнт регресії (6.21), (6.24) можна виразити через коефіцієнти кореляції:

$$\rho_{yx} = r \cdot \frac{\widehat{\sigma}_y}{\widehat{\sigma}_x}. \qquad (6.26)$$

Звідси

$$r^2 = \rho_{yx}\rho_{xy} \text{ або } r = \pm\sqrt{\rho_{yx}\rho_{xy}}, \qquad (6.27)$$

тобто, коефіцієнт кореляції змінних X та Y є середнє геометричне спряжених коефіцієнтів регресії. У формулі (6.27) знак «+» береться, якщо коефіцієнти регресії додатні, і «-» – якщо від'ємні.

Основні властивості коефіцієнта кореляції вибірки:

1. Абсолютне значення коефіцієнта кореляції не перевищує одиниці, тобто

$$|r| \le 1 \text{ або } -1 \le r \le 1. \qquad (6.28)$$

2. Чим ближче значення $|r|$ до одиниці, тим сильніше виражена лінійна залежність. Якщо $r = \pm 1$, то X і Y зв'язані лінійною функціональною залежністю. При цьому лінії регресії Y на X, X на Y співпадають і всі значення вибірки належать цій прямій.

Для якісної оцінки сили зв'язку на основі коефіцієнта кореляції можна використати співвідношення Чеддока (табл. 6.6).

Таблиця 6.6

Величина коефіцієнта кореляції	**Тіснота зв'язку**
$0,9 \le \|r\| < 0,99$	дуже тісний
$0,7 \le \|r\| < 0,9$	тісний
$0,5 \le \|r\| < 0,7$	помітний
$0,3 \le \|r\| < 0,5$	помірний
$0,1 \le \|r\| < 0,3$	слабкий

3. Якщо $r = 0$, то між змінними X та Y відсутній лінійний кореляційний зв'язок (*некорельовані* змінні), але це не виключає існування іншої (нелінійної) форми залежності між ними.

Оскільки коефіцієнт кореляції r обчислюється за даними вибірки, то він є випадковою величиною. Нехай обчислене за вибіркою значення коефіцієнта кореляції свідчить про лінійний зв'язок між змінними. Тоді виникає питання, чи існує лінійна кореляційна залежність між змінними X та Y у генеральній сукупності, чи одержаний результат потрібно віднести на випадковість відбору змінних у вибірку. На практиці ця проблема вирішується шляхом перевірки гіпотези про значущість коефіцієнта кореляції, тобто чи суттєво r відрізняється від нуля, чи цю відмінність можна віднести на рахунок випадковості.

Для вирішення цієї проблеми перевіряється гіпотеза $H_0 : \rho = 0$ (між змінними X та Y не існує лінійного зв'язку) за альтернативної гіпотези $H_1 : \rho \neq 0$ (між змінними існує лінійний зв'язок).

При перевірці використовується критерій

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}. \tag{6.29}$$

Випадкова величина (6.29) має t – розподіл Стьюдента з $n-2$ ступенями свободи (за вибіркою обсягу n для знаходження r обчислюються дві величини $\overline{x}, \overline{y}$). За відомим рівнем значущості α із таблиці значень t – розподілу Стьюдента (додаток 6) визначається критична точка $t_{кр} = t(\alpha; n-2)$ двобічної області, а значення t^* обчислюється за формулою (6.29).

Для знаходження статистичних оцінок параметрів регресії та коефіцієнта кореляції використовуються:

- статистична функція **НОРМОБР(*вероятность*;*среднее*;*стандартное_откл*)**, де *вероятность* – відповідне значення ймовірності, *среднее* – середнє арифметичне значення розподілу;
- функції **slope(*x*,*y*)** – повертає скаляр: кутовий коефіцієнт нахилу лінії регресії до осі *OX* (коефіцієнт регресії), **intercept(*x*,*y*)** – повертає скаляр: зміщення лінії регресії відносно початку координат по осі *OX*,

corr(*x*,*y*) – повертає скаляр: коефіцієнт кореляції. Також ці величини можна розрахувати безпосередньо за формулою.

<u>Приклад 6.1.</u>

Залежність місячного окладу від стажу роботи подана у вигляді таблиці:

Оклад, ум.од.	202	204	212	215	218	220	228	232	240	242	252	260
Стаж роботи (роки)	2	1	2	3	3	1	2	5	3	4	4	7

продовження

Оклад, ум.од.	262	265	270	276	282	284	308	313	312	327	345
Стаж роботи (роки)	7	6	5	8	10	9	13	11	12	9	14

Скласти рівняння лінійної регресії X на Y, визначити коефіцієнт кореляції, зв'язок зобразити графічно. При $\alpha = 0,05$ перевірити гіпотезу про значущість коефіцієнта кореляції між змінними X та Y.

<u>Розв'язання.</u>

Нехай X – місячний оклад (ум. од.), а Y – стаж роботи (роки):

$$X = \{x_i\}, Y = \{y_i\}, i = \overline{1,24}.$$

Рівняння лінійної регресії X на Y має вигляд (6.23): $x^* = \rho_{xy}(y - \overline{y}) + \overline{x}$, де $\rho_{xy} = \dfrac{\overline{xy} - \overline{x} \cdot \overline{y}}{\overline{y}^2 - (\overline{y})^2}$ – коефіцієнт регресії X на Y, $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$, $\overline{y} = \dfrac{1}{n}\sum_{i=1}^{n} y_i$, $\overline{xy} = \dfrac{1}{n}\sum_{i=1}^{n} x_i y_i$.

Всі необхідні проміжні обчислення подано у таблиці 6.7:

Таблиця 6.7

№	x_i	y_i	$x_i \cdot y_i$	x_i^2	y_i^2
1	202	2	404	40804	4
2	204	1	204	41616	1
3	212	2	424	44944	4
4	215	3	645	46225	9
5	218	3	654	47524	9
6	220	1	220	48400	1
7	220	4	880	48400	16
8	228	2	456	51984	4
9	232	5	1160	53824	25
10	240	3	720	57600	9
11	242	4	968	58564	16
12	252	4	1008	63504	16
13	260	7	1820	67600	49
14	262	7	1834	68644	49
15	265	6	1590	70225	36
16	270	5	1350	72900	25
17	276	8	2208	76176	64
18	282	10	2820	79524	100
19	284	9	2556	80656	81
20	308	13	4004	94864	169
21	313	11	3443	97969	121
22	312	12	3744	97344	144
23	327	9	2943	106929	81
24	345	14	4830	119025	196
Сума	**6189**	**145**	**40885**	**1635245**	**1229**

Враховуючи проміжні обчислення (табл. 6.7):

$$\overline{x}=\frac{1}{24}\cdot 6189=257,875;\ \ \overline{y}=\frac{1}{24}\cdot 145=6,042;\ \ \overline{xy}=\frac{1}{24}\cdot 40885=1703,542;$$

$$\overline{y^2}=\frac{1}{24}\cdot 1229=51,208;\ \ \overline{x^2}=\frac{1}{24}\cdot 1635245=68135,208.$$

Тоді коефіцієнт регресії X на Y: $\rho_{xy}=\dfrac{1703,542-257,875\cdot 6,042}{51,208-(6,042)^2}\approx 9,897.$

Отже, рівняння регресії X на Y має вигляд

$$x^*=9,897(y-6,042)+257,875 \text{ або } x^*=9,897y+198,077.$$

Коефіцієнт регресії ρ_{xy} характеризує вплив зміни Y на X, він показує на скільки одиниць в середньому зміниться X при зміні Y на 1 одиницю (у даному випадку на 9,897). Оскільки коефіцієнт регресії ρ_{xy} додатній, то спостерігається позитивний зв'язок між ознаками.

Коефіцієнт кореляції знаходиться за формулою (6.27): $r=\pm\sqrt{\rho_{yx}\rho_{xy}}$.

Оскільки $\rho_{yx}=\dfrac{1703,542-257,875\cdot 6,042}{68135,208-(257,875)^2}\approx 0,089,$ то

$$r=\pm\sqrt{9,897\cdot 0,089}\approx 0,94.$$

Коефіцієнт кореляції близький до одиниці, отже, між досліджуваними величинами існує дуже тісна лінійна залежність.

На координатній площині рівняння регресії та дані значення досліджуваних ознак зображені на рис. 6.3:

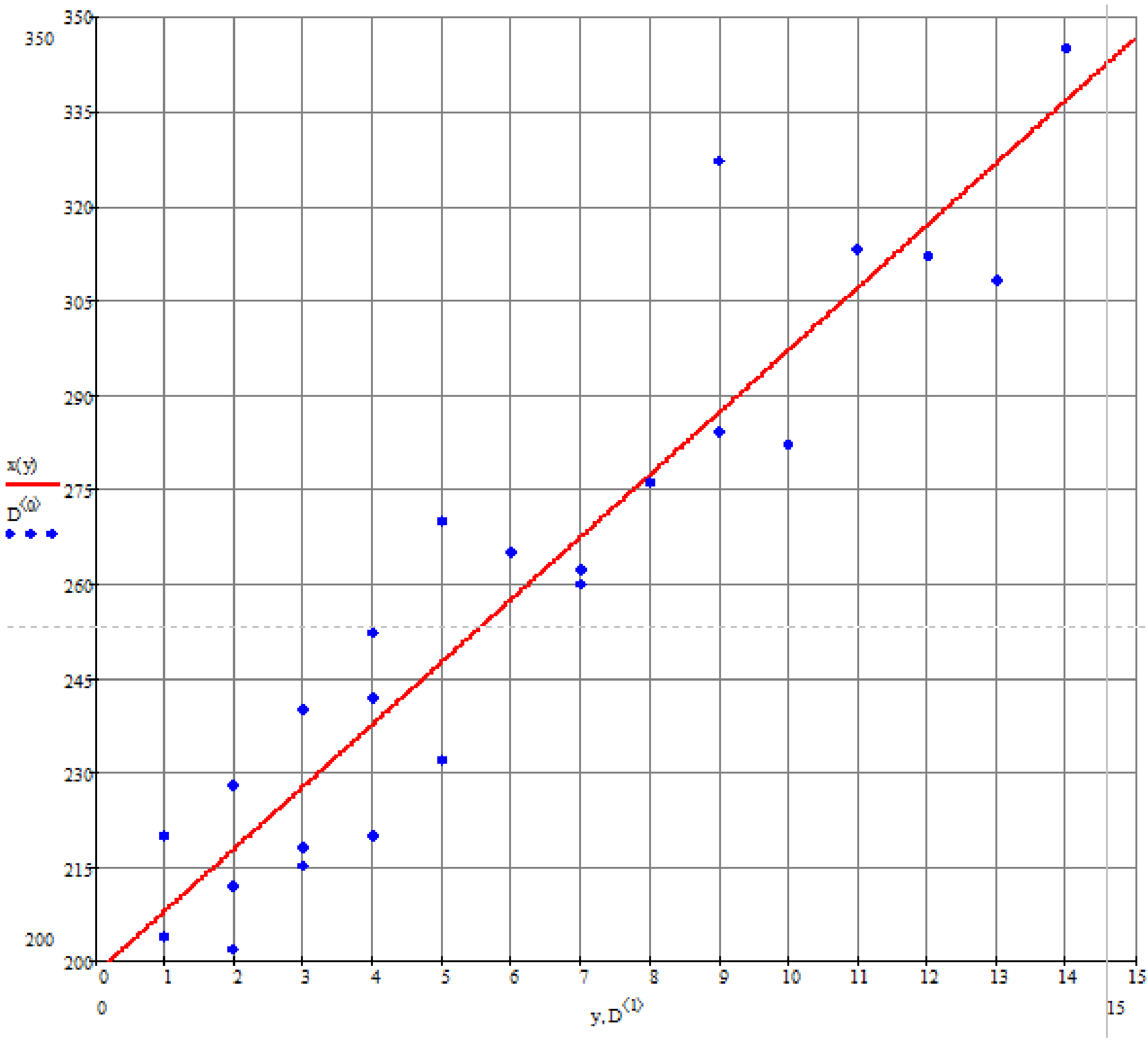


Рис. 6.3

Для перевірки гіпотези про значущість коефіцієнта кореляції між змінними X та Y, слід перевірити гіпотезу $H_0 : \rho = 0$ за альтернативної гіпотези $H_1 : \rho \neq 0$.

За формулою (6.29): $t = \dfrac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ знаходиться спостережуване значення критерію: $t^* = \dfrac{0{,}94 \cdot \sqrt{22}}{\sqrt{1-(0{,}94)^2}} \approx 12{,}9.$

Оскільки $n = 24$, $\alpha = 0,05$, то за таблицею значень розподілу Стьюдента (додаток 5) знаходиться $t_{кр} = t(\alpha; n-2) = t(0,05; 22) = 2,07$ та будується двобічна критична область (рис. 6.4):

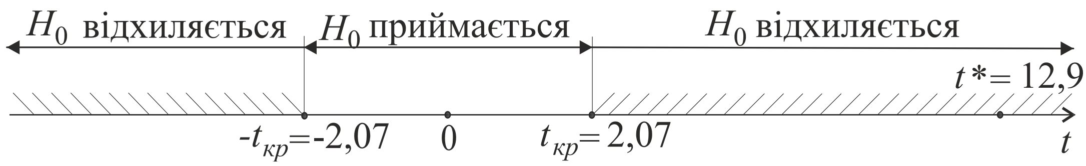


Рис. 6.4

Оскільки $t^* > t_{кр}$ належить критичній області, то при $\alpha = 0,05$ приймається гіпотеза H_1, тобто коефіцієнт кореляції у генеральній сукупності відмінний від нуля, тобто між місячним окладом і стажем роботи існує дуже тісна лінійна кореляційна залежність.

І спосіб

$$XY := \begin{pmatrix} 265 & 6 \\ 282 & 10 \\ 240 & 3 \\ 312 & 12 \\ 220 & 1 \\ 270 & 5 \\ 345 & 14 \\ 212 & 2 \\ 327 & 9 \\ 242 & 4 \\ 218 & 3 \\ 202 & 2 \\ 232 & 5 \end{pmatrix}$$

Примітка. У зв'язку з тим, що матриця *XY* займає багато місця, вона обрізана (її стовпці співпадають із 2 та 3 стовпцями табл. 6.7).

$$ORIGIN := 1$$

$$x := \frac{1}{rows(XY)} \cdot \sum_{i=1}^{rows(XY)} XY_{i,1} = 257.875 \qquad x2 := \frac{1}{rows(XY)} \cdot \sum_{i=1}^{rows(XY)} \left(XY_{i,1}\right)^2 = 6.814 \times 10^4$$

$$y := \frac{1}{rows(XY)} \cdot \sum_{i=1}^{rows(XY)} XY_{i,2} = 6.042 \qquad y2 := \frac{1}{rows(XY)} \cdot \sum_{i=1}^{rows(XY)} \left(XY_{i,2}\right)^2 = 51.208$$

$$xy := \frac{1}{rows(XY)} \cdot \sum_{i=1}^{rows(XY)} \left(XY_{i,1} \cdot XY_{i,2}\right) = 1.704 \times 10^3$$

$$\rho xy := \frac{xy - x \cdot y}{y2 - y^2} = 9.897 \qquad \rho yx := \frac{xy - x \cdot y}{x2 - x^2} = 0.089 \qquad r := \sqrt{\rho xy \cdot \rho yx} = 0.938$$

II спосіб

$$slope\left(XY^{\langle 2 \rangle}, XY^{\langle 1 \rangle}\right) = 9.897 \qquad intercept\left(XY^{\langle 2 \rangle}, XY^{\langle 1 \rangle}\right) = 198.082$$

$$slope\left(XY^{\langle 1 \rangle}, XY^{\langle 2 \rangle}\right) = 0.089 \qquad intercept\left(XY^{\langle 1 \rangle}, XY^{\langle 2 \rangle}\right) = -16.905$$

$$corr\left(XY^{\langle 1 \rangle}, XY^{\langle 2 \rangle}\right) = 0.938$$

Примітка. Рис. 6.3 було скопійовано із програми *MathCad*, тому не є доцільним знову його приводити.

Відповідь: $x^* = 9{,}897y + 198{,}077$; $r \approx 0{,}94$; при $\alpha = 0{,}05$ між місячним окладом і стажем роботи існує дуже тісна лінійна кореляційна залежність.

Практичне заняття 6

6.1. У результаті аналізу інноваційної активності деякого підприємства за певний період часу було отримано залежність між витратами на інновації X (тис. ум. од.) та обсягом реалізованої інноваційної продукції Y (тис. ум. од.):

X	2,1	3,3	1,4	1,8	1,9	2,4	2,7	2	3,1	2,6	3
Y	1,2	1,5	0,7	0,7	0,8	1,4	1,8	1,1	1,3	1,2	1,6

Побудувати кореляційне поле, скласти рівняння лінійної регресії X на Y, визначити коефіцієнт кореляції, зв’язок зобразити графічно. При $\alpha = 0,01$ перевірити гіпотезу про значущість коефіцієнта кореляції між змінними X та Y.

6.2. За даними кореляційної таблиці:

X	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5
Y	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7

побудувати кореляційне поле, скласти і зобразити графічно рівняння лінійної регресії X на Y та Y на X, оцінити тісноту зв’язку між змінними X та Y. Оцінити значущість коефіцієнта кореляції при $\alpha = 0,05$.

6.3. Залежність між витратами на рекламу X (тис. ум. од.) та прибутком Y (тис. ум. од.) деякої мережі магазинів задано у вигляді таблиці:

X	1	2	3	4	5	6	7
Y	1,2	1,7	1,8	2	2,5	2,7	2,8

Побудувати кореляційне поле, скласти і зобразити графічно рівняння лінійної регресії X на Y та Y на X, визначити коефіцієнт кореляції. Оцінити значущість коефіцієнта кореляції при $\alpha = 0,01$.

6.4. Оцінити тісноту зв’язку та значущість $(\alpha = 0,05)$ урожайності соняшника X (т/га) і вологості ґрунту Y (мм), використовуючи дані:

X	3	2,67	2,95	2,34	3,15	2,68	2,98	2,71	3,3	2,22
Y	162	163	160	155	163	158	161	163	165	154

6.5. Оцінити тісноту зв'язку та значущість $(\alpha = 0,01)$ вмісту жиру X (%) і вміст білка Y (%) у насінні соняшнику, використовуючи дані:

X	51,1	47,2	46,5	47,1	47,8	44,8	43,5	47,2	45,4	46,3
Y	13,4	14,4	15,4	14,4	16,2	16,0	15,9	16,2	15,9	16,3

6.6. За даними кореляційної таблиці:

X \ Y	10	20	30	40
0,5	5	1	-	-
1,5	2	5	3	-
2,5	-	7	6	-
3,5	-	-	5	2
4,5	-	-	1	1

побудувати кореляційне поле, скласти і зобразити графічно рівняння лінійної регресії X на Y та Y на X, оцінити тісноту зв'язку між змінними X та Y. Оцінити значущість коефіцієнта кореляції при $\alpha = 0,05$.

6.7. Залежність між ціною товару X (ум. од.) та попитом на нього Y (тис. од.) подано у таблиці:

X \ Y	1	2	3	4
1,5	1	-	-	-
2,5	2	3	-	-
3,5	-	6	7	-
4,5	-	7	5	4
5,5	-	-	-	2

Побудувати кореляційне поле, скласти і зобразити графічно рівняння лінійної регресії X на Y та Y на X, визначити коефіцієнт кореляції. Оцінити значущість коефіцієнта кореляції при $\alpha = 0,05$.

6.8. Для дослідження впливу обсягу капіталовкладень X (млн. ум. од.) на щорічний прибуток Y (млн. ум. од.) було проаналізовано дані 25 підприємств певного регіону:

X \ Y	0-5	5-10	10-15	15-20	20-25
1-2	1	-	-	-	-
2-3	2	3	4	-	-
3-4	-	4	6	3	-
4-5	-	-	-	-	2

Побудувати кореляційне поле, скласти і зобразити графічно рівняння лінійної регресії X на Y та Y на X, визначити коефіцієнт кореляції. Оцінити значущість коефіцієнта кореляції при $\alpha = 0,01$.

6.9. Для аналізу залежності ціни автомобіля X (тис. ум. од.) від пробігу Y (тис. км) було проаналізовано дані певного регіону:

X \ Y	0-25	25-50	50-75	75-100
1-3	-	-	-	1
3-5	-	3	5	2
5-7	2	7	7	-
7-9	5	-	-	-

Побудувати кореляційне поле, скласти і зобразити графічно рівняння лінійної регресії X на Y та Y на X, визначити коефіцієнт кореляції. Оцінити значущість коефіцієнта кореляції при $\alpha = 0,01$.

6.10. Залежність між обсягом випуску продукції X (тис. од.) та її собівартістю Y (ум. од.) подано у таблиці:

X \ Y	1-1,5	1,5-2,0	2,0-2,5	2,5-3
0-5	-	-	1	3
5-10	-	3	4	2
10-15	-	5	2	-
15-20	3	4	-	-
20-25	7	-	-	-

Побудувати кореляційне поле, скласти і зобразити графічно рівняння лінійної регресії X на Y та Y на X, визначити коефіцієнт кореляції. Оцінити значущість коефіцієнта кореляції при $\alpha = 0,05$.

Зразки самостійних та контрольних робіт

Самостійна робота
на тему «Вибірки та їх представлення. Основні числові характеристики статистичного розподілу вибірки»

Варіант 1

За заданим інтервальним статистичним розподілом вибірки

$[\alpha_{i-1};\alpha_i)$	0-1	1-2	2-3
n_i	4	8	6

Потрібно: 1) побудувати гістограму частот; 2) побудувати емпіричну функцію розподілу; 3) знайти дискретний статистичний розподіл частот і відносних частот; 4) побудувати полігон відносних частот; 5) знайти числові характеристики вибірки $(M(X), D(X), \sigma(X))$.

Варіант 2

За заданим інтервальним статистичним розподілом вибірки

$[\alpha_{i-1};\alpha_i)$	1,5-3,5	3,5-5,5	5,5-7,5
n_i	2	3	3

Потрібно: 1) побудувати гістограму частот; 2) побудувати емпіричну функцію розподілу; 3) знайти дискретний статистичний розподіл частот і відносних частот; 4) побудувати полігон відносних частот; 5) знайти числові характеристики вибірки $(M(X), D(X), \sigma(X))$.

Залікова контрольна робота

з математичної статистики

Варіант 1

1. На підприємстві внаслідок впровадження в поточному році нової техніки проводиться обстеження підвищення продуктивності праці (відносно попереднього року). З цією метою вивчаються показники праці 100 робітників. Встановлено, що у 42 продуктивність підвищилась на 4-5 %, у 26 – на 3-4 %, у 18 – на 5-6 %, у 10 – на 2-3 % і в 4 – на 6-7 %. Побудувати: 1) статистичний розподіл вибірки; 2) гістограму відносних частот. Знайти варіаційний розмах.

2. Підприємство випускає прохолоджувальні напої. Тривалий час випадкова величина X – вміст напоїв у пляшці, відповідала нормальному розподілу із стандартним відхиленням 2 мл. Для контролю роботи наповнювального автомата щозміни довільним чином відбираються 64 пляшки з напоями. Результати перевірки вмісту наведені в таблиці:

Об'єм, мл	297	298	299	300	301
К-сть пляшок	2	8	15	25	14

Потрібно: 1) обчислити $\overline{x}_в$, $\sigma_в$; 2) з надійністю $\gamma = 0{,}99$ визначити надійний інтервал для оцінки $M(X)$.

3. Зовнішній аудитор перевіряє систему обліку малого підприємства. Перша перевірка 100 угод показала, що 65 із них помилкові. Підприємству надали місяць на виправлення недоліків. Після цього аудитор провів другу перевірку і виявив 38 угод з порушенням серед 88 перевірених. Чи є підстави вважати, що при рівні значущості $\alpha = 0{,}025$ частка порушень зменшилась між перевірками?

Варіант 2

1. Обстежені головні фонди (млн.грн.) 40 підприємств:

2,8	8,0	2,7	2,4	3,3	6,8	4,3	6,5
4,7	4,8	7,3	7,5	4,6	8,2	5,6	5,5
6,6	2,2	5,9	5,8	7,1	2,3	3,1	7,3
2,3	3,5	3,7	4,9	3,6	2,3	6,5	4,8
4,7	2,2	5,2	3,9	2,7	4,7	5,9	3,7

Побудувати: 1) інтервальний розподіл з інтервалом 1 млн. грн.; 2) кумулятивну криву.

2. Досліджується величина позик банку. З цією метою навмання відібрано 25 клієнтів, які вперше брали гроші під заставу протягом останнього року. Результати аналізу наведені в таблиці:

Позика, тис.грн.	50	80	100	150	200
Кількість клієнтів, чол.	2	5	6	8	4

Вважаючи, що випадкова величина X – сума позики, розподілена за нормальним законом, потрібно: 1) обчислити точкові незсунені статистичні оцінки для $M[X]$, $\sigma[X]$; 2) з надійністю $\gamma = 0,99$ визначити надійний інтервал $\sigma[X]$.

3. Існує два різні способи серійного виробництва підприємством фруктового пюре. Вирішено використовувати той спосіб, який дає більший вміст пектину. Перевірено 10 та 12 виробів, виготовлених відповідно першим та другим способом і знайдено вибіркові дисперсії $D_{в_1} = 15$ кв. од. і $D_{в_2} = 11,5$ кв. од. При рівні значущості $\alpha = 0,01$ перевірити достовірність статистичної гіпотез $H_0 : D_{в_1} = D_{в_2}$, $H_1 : D_{в_1} > D_{в_2}$.

ПРЕДМЕТНИЙ ПОКАЖЧИК

СПИСОК ЛІТЕРАТУРИ

1. Вентцель Е. С. Теория вероятностей : учеб. для вузов / Е. С. Вентцель. – 5-е изд., стер. – М. : Высш. шк., 1998. – 576 с.

2. Берков Н. А. Применением пакета Mathcad: Математический практикум / Н. А. Берков, Н. Н. Елисеева – М. : МГИУ, 2006. – 135 с.

3. Гмурман В. Е. Теория вероятностей и математическая статистика : учеб. для вузов / В. Е. Гмурман. – 6-е изд., стер. – М. : Высш. шк., 1997. – 479 с.

4. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике : учеб. пособие для вузов / В. Е. Гмурман. – 5-е изд., стер. – М. : Высш. шк., 2000. – 400 с.

5. Гнеденко Б. В. Курс теории вероятностей : Учебник. 7-е изд., исправл. – М. : Эдиториал УРСС, 2001. – 320 с.

6. Гурский Д. А. Вычисления в Mathcad 12 / Д. А. Гурский, Е. С. Турбина – СПб. : Питер, 2006. – 544 с.

7. Кремер Н. Ш. Теория вероятностей и математическая статистика : Учебник для вузов / Н. Ш. Кремер. – М. : ЮНИТИ-ДАТА, 2004. – 573 с.

8. Математическая статистика (с примерами в Excel) : учеб. пособие / Ю. Е. Воскобойников, Е. И. Тимошенко ; Новосиб. гос. архитектур.-строит. ун-т (Сибстрин). – 2-е изд., перераб. и доп. – Новосибирск : НГА-СУ (Сибстрин), 2006. – 152 с.

9. Макарова Н. В. Статистика в Excel: Учеб. Пособие / Н. В. Макарова, В. Я. Трофимец. – М. : Финансы и статистика, 2002. – 368 с.

10. Мартиненко М. А. Математична статистика. Навчальний посібник для студентів вищих навчальних закладів / М. А. Мартиненко, О. М. Нещадим, О. І. Радзієвська, В. М. Сафонов. – К. : «Четверта хвиля», 2005. – 257 с.

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

ДОДАТКИ

Додаток 1. Таблиця значень функції Гаусса $\varphi(x)=\frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

	0	1	2	3	4	5	6	7	8	9
0,0	0,3989	3989	3989	3988	3986	3984	3982	3980	3977	3973
0,1	3970	3965	3961	3956	3951	3945	3939	3932	3925	3918
0,2	3910	3902	3894	3885	3876	3867	3857	3847	3836	3825
0,3	3814	3802	3790	3778	3765	3752	3739	3726	3712	3697
0,4	3683	3668	3652	3637	3621	3605	3589	3572	3555	3538
0,5	3521	3503	3485	3467	3448	3429	3410	3391	3372	3352
0,6	3332	3312	3292	3271	3251	3230	3209	3187	3166	3144
0,7	3123	3101	3079	3056	3034	3011	2989	2966	2943	2920
0,8	2897	2874	2850	2827	2803	2780	2756	2732	2709	2685
0,9	2661	2637	2613	2589	2565	2541	2516	2492	2468	2444
1,0	2420	2396	2371	2347	2323	2299	2275	2251	2227	2203
1,1	2179	2155	2131	2107	2083	2059	2036	2012	1989	1965
1,2	1942	1919	1895	1872	1849	1826	1804	1781	1758	1736
1,3	1714	1691	1669	1647	1626	1604	1582	1561	1539	1518
1,4	1497	1476	1456	1435	1415	1394	1374	1354	1334	1315
1,5	1295	1276	1257	1238	1219	1200	1182	1163	1145	1127
1,6	1109	1092	1074	1057	1040	1023	1006	0989	0973	0957
1,7	0940	0925	0909	0893	0878	0863	0848	0833	0818	0804
1,8	0790	0775	0761	0748	0734	0721	0707	0694	0681	0669
1,9	0656	0644	0632	0620	0608	0596	0584	0573	0562	0551
2,0	0540	0529	0519	0508	0498	0488	0478	0468	0459	0449
2,1	0440	0431	0422	0413	0404	0396	0387	0379	0371	0363
2,2	0355	0347	0339	0332	0325	0317	0310	0303	0297	0290
2,3	0283	0277	0270	0264	0258	0252	0246	0241	0235	0229
2,4	0224	0219	0213	0208	0203	0198	0194	0189	0184	0180
2,5	0175	0171	0167	0163	0158	0154	0151	0147	0143	0139
2,6	0136	0132	0129	0126	0122	0119	0116	0113	0110	0107
2,7	0104	0101	0099	0096	0093	0091	0088	0086	0084	0081
2,8	0079	0077	0075	0073	0071	0069	0067	0065	0063	0061
2,9	0060	0058	0056	0055	0053	0051	0050	0048	0047	0046
3,0	0044	0043	0042	0040	0039	0038	0037	0036	0035	0034
3,1	0033	0032	0031	0030	0029	0028	0027	0026	0025	0025
3,2	0024	0023	0022	0022	0021	0020	0020	0019	0018	0018
3,3	0017	0017	0016	0016	0015	0015	0014	0014	0013	0013
3,4	0012	0012	0012	0011	0011	0010	0010	0010	0009	0009
3,5	0009	0008	0008	0008	0008	0007	0007	0007	0007	0006
3,6	0006	0006	0006	0005	0005	0005	0005	0005	0005	0004
3,7	0004	0004	0004	0004	0004	0004	0003	0003	0003	0003
3,8	0003	0003	0003	0003	0003	0002	0002	0002	0002	0002
3,9	0002	0002	0002	0002	0002	0002	0002	0002	0001	0001

Додаток 2. Таблиця значень функції Лапласа $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int\limits_0^x e^{-\frac{z^2}{2}}dz.$

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,00	0,0000	0,52	0,1985	1,04	0,3508	1,56	0,4406	2,16	0,4846
0,01	0,0040	0,53	0,2019	1,05	0,3531	1,57	0,4418	2,18	0,4854
0,02	0,0080	0,54	0,2054	1,06	0,3554	1,58	0,4429	2,20	0,4861
0,03	0,0120	0,55	0,2088	1,07	0,3577	1,59	0,4441	2,22	0,4868
0,04	0,0160	0,56	0,2123	1,08	0,3599	1,60	0,4452	2,24	0,4875
0,05	0,0199	0,57	0,2157	1,09	0,3621	1,61	0,4463	2,26	0,4881
0,06	0,0239	0,58	0,2190	1,10	0,3643	1,62	0,4474	2,28	0,4887
0,07	0,0279	0,59	0,2224	1,11	0,3665	1,63	0,4484	2,30	0,4893
0,08	0,0319	0,60	0,2257	1,12	0,3686	1,64	0,4495	2,32	0,4898
0,09	0,0359	0,61	0,2291	1,13	0,3708	1,65	0,4505	2,34	0,4904
0,10	0,0398	0,62	0,2324	1,14	0,3729	1,66	0,4515	2,36	0,4909
0,11	0,0438	0,63	0,2357	1,15	0,3749	1,67	0,4525	2,38	0,4913
0,12	0,0478	0,64	0,2389	1,16	0,3770	1,68	0,4535	2,40	0,4918
0,13	0,0517	0,65	0,2422	1,17	0,3790	1,69	0,4545	2,42	0,4922
0,14	0,0557	0,66	0,2454	1,18	0,3810	1,70	0,4554	2,44	0,4927
0,15	0,0596	0,67	0,2486	1,19	0,3830	1,71	0,4564	2,46	0,4931
0,16	0,0636	0,68	0,2517	1,20	0,3849	1,72	0,4573	2,48	0,4934
0,17	0,0675	0,69	0,2549	1,21	0,3869	1,73	0,4582	2,50	0,4938
0,18	0,0714	0,70	0,2580	1,22	0,3863	1,74	0,4591	2,52	0,4941
0,19	0,0753	0,71	0,2611	1,23	0,3907	1,75	0,4599	2,54	0,4945
0,20	0,0793	0,72	0,2642	1,24	0,3925	1,76	0,4608	2,56	0,4948
0,21	0,0832	0,73	0,2673	1,25	0,3944	1,77	0,4616	2,58	0,4951
0,22	0,0871	0,74	0,2703	1,26	0,3962	1,78	0,4625	2,60	0,4953
0,23	0,0910	0,75	0,2734	1,27	0,3980	1,79	0,4633	2,62	0,4956
0,24	0,0948	0,76	0,2764	1,28	0,3997	1,80	0,4641	2,64	0,4959
0,25	0,0987	0,77	0,2794	1,29	0,4015	1,81	0,4649	2,66	0,4961
0,26	0,1026	0,78	0,2823	1,30	0,4032	1,82	0,4656	2,68	0,4963
0,27	0,1064	0,79	0,2852	1,31	0,4049	1,83	0,4664	2,70	0,4965
0,28	0,1103	0,80	0,2881	1,32	0,4066	1,84	0,4671	2,72	0,4967
0,29	0,1141	0,81	0,2910	1,33	0,4082	1,85	0,4678	2,74	0,4669
0,30	0,1179	0,82	0,2939	1,34	0,4099	1,86	0,4686	2,76	0,4971
0,31	0,1217	0,83	0,2967	1,35	0,4115	1,87	0,4693	2,78	0,4973
0,32	0,1255	0,84	0,2965	1,36	0,4131	1,88	0,4699	2,80	0,4974
0,33	0,1293	0,85	0,3023	1,37	0,4147	1,89	0,4706	2,82	0,4976
0,34	0,1331	0,86	0,3051	1,38	0,4162	1,90	0,4713	2,84	0,4977
0,35	0,1368	0,87	0,3078	1,39	0,4177	1,91	0,4719	2,86	0,4979
0,36	0,1406	0,88	0,3106	1,40	0,4192	1,92	0,4726	2,88	0,4980
0,37	0,1443	0,89	0,3133	1,41	0,4207	1,93	0,4732	2,90	0,4981
0,38	0,1480	0,90	0,3159	1,42	0,4222	1,94	0,4738	2,92	0,4982
0,39	0,1517	0,91	0,3186	1,43	0,4236	1,95	0,4744	2,94	0,4984
0,40	0,1554	0,92	0,3212	1,44	0,4251	1,96	0,4750	2,96	0,4985
0,41	0,1591	0,93	0,3238	1,45	0,4265	1,97	0,4756	2,98	0,4986
0,42	0,1628	0,94	0,3264	1,46	0,4279	1,98	0,4761	3,00	0,4986
0,43	0,1664	0,95	0,3289	1,47	0,4292	1,99	0,4767	3,20	0,4993
0,44	0,1700	0,96	0,3315	1,48	0,4306	2,00	0,4772	3,40	0,4996
0,45	0,1736	0,97	0,3340	1,49	0,4319	2,02	0,4783	3,60	0,4998

x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$	x	$\Phi(x)$
0,46	0,1772	0,98	0,3365	1,50	0,4332	2,04	0,4793	3,80	0,4999
0,47	0,1808	0,99	0,3389	1,51	0,4345	2,06	0,4803	4,00	0,4999
0,48	0,1844	1,00	0,3413	1,52	0,4357	2,08	0,4812	4,50	0,4999
0,49	0,1879	1,01	0,3438	1,53	0,4370	2,10	0,4821	5,00	0,4999
0,50	0,1915	1,02	0,3461	1,54	0,4382	2,12	0,4830		
0,51	0,1950	1,03	0,3185	1,55	0,4994	2,14	0,4838		

Додаток 3. Таблиця значень функції $f(x) = e^{-x}$.

x	$f(x)$	x	$f(x)$	x	$f(x)$	x	$f(x)$	x	$f(x)$
0,00	1,0000	0,40	0,6703	0,80	0,4493	1,20	0,3012	1,60	0,2019
0,01	0,9900	0,41	0,6637	0,81	0,4449	1,21	0,2982	1,61	0,1999
0,02	0,9802	0,42	0,6570	0,82	0,4404	1,22	0,2952	1,62	0,1979
0,03	0,9704	0,43	0,6505	0,83	0,4360	1,23	0,2923	1,63	0,1959
0,04	0,9608	0,44	0,6440	0,84	0,4317	1,24	0,2894	1,64	0,1940
0,05	0,9512	0,45	0,6376	0,85	0,4274	1,25	0,2865	1,65	0,1920
0,06	0,9418	0,46	0,6313	0,86	0,4232	1,26	0,2837	1,66	0,1901
0,07	0,9324	0,47	0,6250	0,87	0,4190	1,27	0,2808	1,67	0,1882
0,08	0,9231	0,48	0,6188	0,88	0,4148	1,28	0,2780	1,68	0,1864
0,09	0,9139	0,49	0,6126	0,89	0,4107	1,29	0,2753	1,69	0,1845
0,10	0,9048	0,50	0,6065	0,90	0,4066	1,30	0,2725	1,70	0,1827
0,11	0,8958	0,51	0,6005	0,91	0,4025	1,31	0,2698	1,71	0,1809
0,12	0,8869	0,52	0,5945	0,92	0,3985	1,32	0,2671	1,72	0,1791
0,13	0,8781	0,53	0,5886	0,93	0,3946	1,33	0,2645	1,73	0,1773
0,14	0,8694	0,54	0,5827	0,94	0,3906	1,34	0,2618	1,74	0,1755
0,15	0,8607	0,55	0,5769	0,95	0,3867	1,35	0,2592	1,75	0,1738
0,16	0,8521	0,56	0,5712	0,96	0,3829	1,36	0,2567	1,76	0,1720
0,17	0,8437	0,57	0,5655	0,97	0,3791	1,37	0,2541	1,77	0,1703
0,18	0,8353	0,58	0,5599	0,98	0,3753	1,38	0,2516	1,78	0,1686
0,19	0,8270	0,59	0,5543	0,99	0,3716	1,39	0,2491	1,79	0,1670
0,20	0,8187	0,60	0,5488	1,00	0,3679	1,40	0,2466	1,80	0,1653
0,21	0,8106	0,61	0,5434	1,01	0,3642	1,41	0,2441	1,81	0,1637
0,22	0,8025	0,62	0,5379	1,02	0,3606	1,42	0,2417	1,82	0,1620
0,23	0,7945	0,63	0,5326	1,03	0,3570	1,43	0,2393	1,83	0,1604
0,24	0,7866	0,64	0,5273	1,04	0,3535	1,44	0,2369	1,84	0,1588
0,25	0,7788	0,65	0,5220	1,05	0,3499	1,45	0,2346	1,85	0,1572
0,26	0,7711	0,66	0,5169	1,06	0,3465	1,46	0,2322	1,86	0,1557
0,27	0,7634	0,67	0,5117	1,07	0,3430	1,47	0,2299	1,87	0,1541
0,28	0,7558	0,68	0,5066	1,08	0,3396	1,48	0,2276	1,88	0,1526
0,29	0,7483	0,69	0,5016	1,09	0,3362	1,49	0,2254	1,89	0,1511
0,30	0,7408	0,70	0,4966	1,10	0,3329	1,50	0,2231	1,90	0,1496
0,31	0,7334	0,71	0,4916	1,11	0,3296	1,51	0,2209	1,91	0,1481
0,32	0,7261	0,72	0,4868	1,12	0,3263	1,52	0,2187	1,92	0,1466
0,33	0,7189	0,73	0,4819	1,13	0,3230	1,53	0,2165	1,93	0,1451
0,34	0,7118	0,74	0,4771	1,14	0,3198	1,54	0,2144	1,94	0,1437
0,35	0,7047	0,75	0,4724	1,15	0,3166	1,55	0,2122	1,95	0,1423
0,36	0,6977	0,76	0,4677	1,16	0,3135	1,56	0,2101	1,96	0,1409
0,37	0,6907	0,77	0,4630	1,17	0,3104	1,57	0,2080	1,97	0,1395
0,38	0,6839	0,78	0,4584	1,18	0,3073	1,58	0,2060	1,98	0,1381
0,39	0,6771	0,79	0,4538	1,19	0,3042	1,59	0,2039	1,99	0,1367

Продовження додатку 3

x	$f(x)$	x	$f(x)$	x	$f(x)$	x	$f(x)$	x	$f(x)$
2,00	0,1353	2,49	0,0829	2,98	0,0508	3,47	0,0311	3,96	0,0191
2,01	0,1340	2,50	0,0821	2,99	0,0503	3,48	0,0308	3,97	0,0189
2,02	0,1327	2,51	0,0813	3,00	0,0498	3,49	0,0305	3,98	0,0187
2,03	0,1313	2,52	0,0805	3,01	0,0493	3,50	0,0302	3,99	0,0185
2,04	0,1300	2,53	0,0797	3,02	0,0488	3,51	0,0299	4,00	0,0183
2,05	0,1287	2,54	0,0789	3,03	0,0483	3,52	0,0296	4,01	0,0181
2,06	0,1275	2,55	0,0781	3,04	0,0478	3,53	0,0293	4,02	0,0180
2,07	0,1262	2,56	0,0773	3,05	0,0474	3,54	0,0290	4,03	0,0178
2,08	0,1249	2,57	0,0765	3,06	0,0469	3,55	0,0287	4,04	0,0176
2,09	0,1237	2,58	0,0758	3,07	0,0464	3,56	0,0284	4,05	0,0174
2,10	0,1225	2,59	0,0750	3,08	0,0460	3,57	0,0282	4,06	0,0172
2,11	0,1212	2,60	0,0743	3,09	0,0455	3,58	0,0279	4,07	0,0171
2,12	0,1200	2,61	0,0735	3,10	0,0450	3,59	0,0276	4,08	0,0169
2,13	0,1188	2,62	0,0728	3,11	0,0446	3,60	0,0273	4,09	0,0167
2,14	0,1177	2,63	0,0721	3,12	0,0442	3,61	0,0271	4,10	0,0166
2,15	0,1165	2,64	0,0714	3,13	0,0437	3,62	0,0268	4,11	0,0164
2,16	0,1153	2,65	0,0707	3,14	0,0433	3,63	0,0265	4,12	0,0162
2,17	0,1142	2,66	0,0699	3,15	0,0429	3,64	0,0263	4,13	0,0161
2,18	0,1130	2,67	0,0693	3,16	0,0424	3,65	0,0260	4,14	0,0159
2,19	0,1119	2,68	0,0686	3,17	0,0420	3,66	0,0257	4,15	0,0158
2,20	0,1108	2,69	0,0679	3,18	0,0416	3,67	0,0255	4,16	0,0156
2,21	0,1097	2,70	0,0672	3,19	0,0412	3,68	0,0252	4,17	0,0155
2,22	0,1086	2,71	0,0665	3,20	0,0408	3,69	0,0250	4,18	0,0153
2,23	0,1075	2,72	0,0659	3,21	0,0404	3,70	0,0247	4,19	0,0151
2,24	0,1065	2,73	0,0652	3,22	0,0400	3,71	0,0245	4,20	0,0150
2,25	0,1054	2,74	0,0646	3,23	0,0396	3,72	0,0242	4,21	0,0148
2,26	0,1044	2,75	0,0639	3,24	0,0392	3,73	0,0240	4,22	0,0147
2,27	0,1033	2,76	0,0633	3,25	0,0388	3,74	0,0238	4,23	0,0146
2,28	0,1023	2,77	0,0627	3,26	0,0384	3,75	0,0235	4,24	0,0144
2,29	0,1013	2,78	0,0620	3,27	0,0380	3,76	0,0233	4,25	0,0143
2,30	0,1003	2,79	0,0614	3,28	0,0376	3,77	0,0231	4,26	0,0141
2,31	0,0993	2,80	0,0608	3,29	0,0373	3,78	0,0228	4,27	0,0140
2,32	0,0983	2,81	0,0602	3,30	0,0369	3,79	0,0226	4,28	0,0138
2,33	0,0973	2,82	0,0596	3,31	0,0365	3,80	0,0224	4,29	0,0137
2,34	0,0963	2,83	0,0590	3,32	0,0362	3,81	0,0221	4,30	0,0136
2,35	0,0954	2,84	0,0584	3,33	0,0358	3,82	0,0219	4,31	0,0134
2,36	0,0944	2,85	0,0578	3,34	0,0354	3,83	0,0217	4,32	0,0133
2,37	0,0935	2,86	0,0573	3,35	0,0351	3,84	0,0215	4,33	0,0132
2,38	0,0926	2,87	0,0567	3,36	0,0347	3,85	0,0213	4,34	0,0130
2,39	0,0916	2,88	0,0561	3,37	0,0344	3,86	0,0211	4,35	0,0129
2,40	0,0907	2,89	0,0556	3,38	0,0340	3,87	0,0209	4,36	0,0128
2,41	0,0898	2,90	0,0550	3,39	0,0337	3,88	0,0207	4,37	0,0127
2,42	0,0889	2,91	0,0545	3,40	0,0334	3,89	0,0204	4,38	0,0125
2,43	0,0880	2,92	0,0539	3,41	0,0330	3,90	0,0202	4,39	0,0124
2,44	0,0872	2,93	0,0534	3,42	0,0327	3,91	0,0200	4,40	0,0123
2,45	0,0863	2,94	0,0529	3,43	0,0324	3,92	0,0198	4,41	0,0122
2,46	0,0854	2,95	0,0523	3,44	0,0321	3,93	0,0196	4,42	0,0120
2,47	0,0846	2,96	0,0518	3,45	0,0317	3,94	0,0194	4,43	0,0119
2,48	0,0837	2,97	0,0513	3,46	0,0314	3,95	0,0193	4,44	0,0118

x	$f(x)$	x	$f(x)$	x	$f(x)$	x	$f(x)$	x	$f(x)$
4,45	0,0117	4,82	0,0081	5,19	0,0056	5,61	0,0037	6,30	0,0018
4,46	0,0116	4,83	0,0080	5,20	0,0055	5,62	0,0036	6,34	0,0018
4,47	0,0114	4,84	0,0079	5,21	0,0055	5,64	0,0036	6,35	0,0017
4,48	0,0113	4,85	0,0078	5,22	0,0054	5,65	0,0035	6,40	0,0017
4,49	0,0112	4,86	0,0078	5,23	0,0054	5,66	0,0035	6,41	0,0016
4,50	0,0111	4,87	0,0077	5,24	0,0053	5,67	0,0034	6,46	0,0016
4,51	0,0110	4,88	0,0076	5,25	0,0052	5,69	0,0034	6,47	0,0015
4,52	0,0109	4,89	0,0075	5,26	0,0052	5,70	0,0033	6,53	0,0015
4,53	0,0108	4,90	0,0074	5,27	0,0051	5,72	0,0033	6,54	0,0014
4,54	0,0107	4,91	0,0074	5,28	0,0051	5,73	0,0032	6,60	0,0014
4,55	0,0106	4,92	0,0073	5,29	0,0050	5,76	0,0032	6,61	0,0013
4,56	0,0105	4,93	0,0072	5,30	0,0050	5,77	0,0031	6,68	0,0013
4,57	0,0104	4,94	0,0072	5,31	0,0049	5,79	0,0031	6,69	0,0012
4,58	0,0103	4,95	0,0071	5,32	0,0049	5,80	0,0030	6,76	0,0012
4,59	0,0102	4,96	0,0070	5,33	0,0048	5,82	0,0030	6,77	0,0011
4,60	0,0101	4,97	0,0069	5,34	0,0048	5,83	0,0029	6,85	0,0011
4,61	0,0100	4,98	0,0069	5,35	0,0047	5,86	0,0029	6,86	0,0010
4,62	0,0099	4,99	0,0068	5,37	0,0047	5,87	0,0028	6,95	0,0010
4,63	0,0098	5,00	0,0067	5,38	0,0046	5,89	0,0028	6,96	0,0009
4,64	0,0097	5,01	0,0067	5,39	0,0046	5,90	0,0027	7,07	0,0009
4,65	0,0096	5,02	0,0066	5,40	0,0045	5,93	0,0027	7,08	0,0008
4,66	0,0095	5,03	0,0065	5,41	0,0045	5,94	0,0026	7,19	0,0008
4,67	0,0094	5,04	0,0065	5,42	0,0044	5,97	0,0026	7,20	0,0007
4,68	0,0093	5,05	0,0064	5,43	0,0044	5,98	0,0025	7,33	0,0007
4,69	0,0092	5,06	0,0063	5,44	0,0043	6,01	0,0025	7,34	0,0006
4,70	0,0091	5,07	0,0063	5,46	0,0043	6,02	0,0024	7,50	0,0006
4,71	0,0090	5,08	0,0062	5,47	0,0042	6,05	0,0024	7,51	0,0005
4,72	0,0089	5,09	0,0062	5,48	0,0042	6,06	0,0023	7,70	0,0005
4,73	0,0088	5,10	0,0061	5,49	0,0041	6,09	0,0023	7,71	0,0004
4,74	0,0087	5,11	0,0060	5,50	0,0041	6,10	0,0022	7,95	0,0004
4,75	0,0087	5,12	0,0060	5,51	0,0040	6,14	0,0022	7,96	0,0003
4,76	0,0086	5,13	0,0059	5,53	0,0040	6,15	0,0021	8,29	0,0003
4,77	0,0085	5,14	0,0059	5,54	0,0039	6,18	0,0021	8,30	0,0002
4,78	0,0084	5,15	0,0058	5,55	0,0039	6,19	0,0020	8,80	0,0002
4,79	0,0083	5,16	0,0057	5,56	0,0038	6,23	0,0020	8,81	0,0001
4,80	0,0082	5,17	0,0057	5,58	0,0038	6,24	0,0019	9,90	0,0001
4,81	0,0081	5,18	0,0056	5,59	0,0037	6,29	0,0019	9,91	0,0000

Додаток 4. Таблиця значень $t_\gamma = t(\gamma, k)$.

$k = n-1$	γ					
	0,8	**0,9**	**0,95**	**0,98**	**0,99**	**0,999**
1	3,078	6,314	12,706	31,821	63,657	63,662
2	1,886	2,920	4,303	6,965	9,925	31,598
3	2,638	2,353	3,182	4,541	5,841	12,941
4	1,533	2,132	2,776	3,747	4,694	8,610
5	1,476	2,015	2,571	3,365	4,032	6,859
6	1,440	1,943	2,447	3,143	3,707	5,959
7	1,415	1,895	2,365	2,998	3,499	5,405
8	1,397	1,860	2,306	2,896	3,355	5,041
9	1,383	1,833	2,262	2,821	3,250	4,781
10	1,372	1,812	2,228	2,764	3,169	4,587
11	1,363	1,796	2,201	2,718	3,106	4,487
12	1,356	1,782	2,179	2,681	3,055	4,318
13	1,350	1,771	2,160	2,650	3,012	4,221
14	1,345	1,761	2,145	2,624	2,977	4,140
15	1,341	1,753	2,131	2,602	2,947	4,073
16	1,337	1,746	2,120	2,583	2,921	4,015
17	1.333	1,740	2,110	2,567	2,898	3,965
18	1,330	1,734	2,103	2,552	2,872	3,922
19	1,328	1,729	2,093	2,539	2,861	3,883
20	1,325	1,725	2.086	2,528	2,845	3,850
21	1,323	1,721	2,080	2,518	2,831	3,819
22	1,321	1,717	2,074	2,508	2,819	3,792
23	1,319	1,714	2,069	2,500	2,807	3,767
24	1,318	1,711	2,064	2,492	2,797	3,745
25	1,316	1,708	2,060	2,485	2,787	3,725
26	1,315	1,706	2,056	2,479	2,779	3,707
27	1,314	1,703	2,052	2,473	2,771	3,690
28	1,313	1,701	2,048	2,467	2,763	3,674
29	1,311	1,699	2,045	2,462	2,756	3,659
30	1,310	1,697	2,042	2,457	2,750	3,646

Додаток 5. Таблиця значень $q = q(\gamma, k)$.

k	γ				
	0,9	**0,95**	**0,98**	**0,99**	**0,999**
3	1,270	1,932	3,000	4,200	9,00
4	0,941	1,382	2,056	2,700	5,00
5	0,738	1,104	1,594	2,000	3,80
6	0,623	0,918	1,306	1,650	3,00
7	0,576	0,800	1,143	1,393	2,50
8	0,516	0,713	0,986	1,225	2,05
9	0,476	0,650	0,889	1,094	1,75
10	0,442	0,596	0,814	0,980	1,50
12	0,338	0,527	0,700	0,840	1,30
14	0,357	0,468	0,620	0,740	1,14
16	0,325	0,422	0,564	0,671	1,02
18	0,297	0,390	0,500	0,600	0,92
20	0,282	0,370	0,480	0,567	0,85
25	0,247	0,317	0,408	0,485	0,70
30	0,226	0,281	0,369	0,425	0,60
35	0,207	0,261	0,347	0,400	0,56
40	0,193	0,242	0,312	0,375	0,52
45	0,184	0,228	0,288	0,350	0,48
50	0,174	0,212	0,270	0,311	0,45
60	0,155	0,193	0,242	0,283	0,40
100	0,125	0,146	0,184	0,200	0,30
1000	0,044	0,047	0,056	0,059	0,08

Додаток 6. Критичні точки $t-$ розподілу Стьюдента

Число ступенів свободи, k	**Рівень значущості** α						
	0,20	0,10	0,05	0,02	0,01	0,002	0,001
1	3,08	6,31	12,7	31,82	63,66	127,32	636,62
2	1,89	2,92	4,30	6,97	9,93	14,09	3131,60
3	1,64	2,35	3,18	4,54	5,84	7,45	12,94
4	1,53	2,13	2,78	3,75	4,60	5,60	8,61
5	1,48	2,02	2,57	3,37	4,03	4,77	6,86
6	1,44	1,94	2,45	3,14	3,71	4,32	5,96
7	1,42	1,90	2,36	3,00	3,50	4,03	5,41
8	1,40	1,86	2,31	2,90	3,36	3,83	5,04
9	1,38	1,83	2,26	2,82	3,25	3,69	4,78
10	1,37	1,81	2,23	2,76	3,17	3,58	4,59
11	1,36	1,80	2,20	2,72	3,11	3,50	4,44
12	1,36	1,78	2,18	2,68	3,05	3,43	4,32
13	1,35	1,77	2,16	2,65	3,01	3,37	4,22
14	1,34	1,76	2,14	2,62	2,98	3,33	4,14
15	1,34	1,75	2,13	2,60	2,95	3,29	4,07
16	1,34	1,75	2,12	2,58	2,92	3,25	4,02
17	1,33	1,74	2,11	2,57	2,90	3,22	3,97
18	1,33	1,73	2,10	2,55	2,88	3,20	3,92
19	1,33	1,73	2,09	2,54	2,86	3,17	3,88

Закінчення додатку 6

Число ступенів свободи, k	Рівень значущості α						
	0,20	**0,10**	**0,05**	**0,02**	**0,01**	**0,002**	**0,001**
20	1,33	1,73	2,09	2,53	2,85	3,15	3,85
21	1,32	1,72	2,08	2,52	2,83	3,14	3,82
22	1,32	1,72	2,07	2,51	2,82	3,12	3,79
23	1,32	1,71	2,07	2,50	2,81	3,10	3,77
24	1,32	1,71	2,06	2,49	2,80	3,09	3,75
25	1,32	1,71	2,06	2,48	2,79	3,08	3,73
26	1,32	1,71	2,06	2,48	2,78	3,07	3,71
27	1,31	1,70	2,05	2,47	2,77	3,06	3,69
28	1,31	1,70	2,05	2,47	2,76	3,05	3,67
29	1,31	1,70	2,04	2,46	2,76	3,04	3,66
30	1,31	1,70	2,04	2,46	2,75	3,03	3,65
40	1,30	1,68	2,02	2,42	2,70	2,97	3,55
60	1,30	1,67	2,00	2,39	2,66	2,91	3,46
120	1,29	1,66	1,98	2,36	2,62	2,86	3,37
∞	1,28	1,64	1,96	2,33	2,58	2,81	3,29

Додаток 7. Критичні точки F-розподілу Фішера-Снедекора

(k_1 - число ступенів свободи більшої дисперсії,

k_2 - число ступенів свободи меншої дисперсії).

k_1 / k_2	Рівень значущості $\alpha = 0,01$											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
1	4052	4999	5403	5625	5674	5889	5928	5981	6022	6056	6082	6106
2	98,49	99,01	90,17	99,25	99,33	99,30	99,34	99,36	99,36	99,40	99,41	99,42
3	34,12	30,81	29,46	28,71	28,24	27,91	27,67	27,49	27,34	27,23	27,13	27,05
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14.54	14,45	14,37
5	16,26	13,27	12,06	11,39	10,97	10,67	10,45	10,27	10,15	10,05	9,96	9,89
6	13,74	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,79	7,72
7	12,25	9,55	8,45	7,85	7,46	7,19	7,00	6,84	6,71	6.62	6,54	6,47
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,82	5,74	5,67
9	10,56	8,02	6,99	6,42	6,06	5,80	5,62	5,47	5,35	5,26	5,18	5,11
10	10,04	7,56	6,55	5,99	5,64	5,39	5,21	5,06	4,95	4,85	4,78	4,71
11	9,86	7,20	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,46	4,10
12	9,33	6,93	5,95	5,41	5,06	4,82	4,65	4,50	4,39	4,30	4,22	4,16
13	9,07	6,70	5,74	5,20	4,86	4,62	4,44	4,30	4,19	4,10	4,02	3,96
14	8,86	6,51	5,56	5,03	4,69	4,46	4,28	4,14	4,03	3,94	3,86	3,80
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,73	3,67
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,61	3,55
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,52	3,45

Рівень значущості $\alpha = 0,05$

1	161	200	216	225	230	234	237	239	241	242	243	244
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,40	19,41
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,76	8,74
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6.04	6,00	5,96	5,93	5,91
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,70	4,68
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,03	4,00

k_2 \ k_1	Рівень значущості $\alpha = 0,05$											
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,60	3,57
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,31	3,28
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,10	3,07
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,94	2,91
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,82	2,79
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,72	2,69
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,63	2,60
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,56	2,53
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,51	2,48
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,45	2,42
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,41	2,38

Додаток 8. Критичні точки розподілу χ^2 (xi - квадрат).

Число ступенів свободи, k	Рівень значущості α					
	0,01	**0,025**	**0,05**	**0,95**	**0,975**	**0,99**
1	6,6	5,0	3,8	0,0039	0,00098	0,00016
2	9,2	7,4	6,0	0,103	0,051	0,020
3	11,3	9,4	7,8	0,352	0,216	0,115
4	13,3	11,1	9,5	0,711	0,484	0,297
5	15,1	12,8	11,1	1,15	0,831	0,554
6	16,8	14,4	12,6	1,64	1,24	0,872
7	18,5	16,0	14,1	2,17	1,69	1,24
8	20,1	17.5	15,5	2,73	2,18	1,65
9	21,7	19,0	16,9	3,33	2,70	2,09
10	23,2	20,5	18,3	3,94	3,25	2,56
11	24,7	21,9	19,7	4,57	3,82	3,05
12	26,2	23,3	21,0	5,23	4,40	3,57
13	27,7	24,7	22,4	5,89	5,01	4,11
14	29,1	26,1	23,7	6,57	5,63	4,66
15	30,6	27,5	25,0	7,26	6,26	5,23
16	32,0	28,8	26,3	7,96	6,91	5,81
17	33,4	30.2	27,6	8,67	7,56	6,41
18	34,8	31,5	28,9	9,39	8,23	7,01
19	36,2	32,9	30,1	10,1	8,91	7,63
20	37,6	34,2	31,4	10,9	9,59	8,26
21	38,9	35,5	32,7	11,6	10,3	8,90
22	40,3	36,8	33,9	12,3	11,0	9,54
23	41,6	38,1	35,2	13,1	11,7	10,2
24	43,0	39,4	36,4	13,8	12,4	10,9
25	44,3	40,6	37,7	14,6	13,1	11,5
26	45,6	41,9	38,9	15,4	13,8	12,2
27	47,0	43,2	40,1	16,2	14,6	12,9
28	48,3	44,5	41,3	16,9	15,3	13,6
29	49,6	45,7	42,6	17,7	16,0	14,3
30	50,9	47,0	43,8	18,5	16,8	15,0

Додаток 9. Значення розподілу χ_i^2

залежно від ймовірності $P(\chi^2(k) > \chi_i^2) = p.$

k	p					
	0,99	**0,95**	**0,90**	**0,10**	**0,05**	**0,01**
1	0,0002	0,004	0,02	2,71	3,84	6,64
2	0,02	0,10	0,21	4,61	5,99	9,21
3	0,12	0,35	0,58	6,25	7,82	11,34
4	0,30	0,71	1,06	7,78	9,49	13,28
5	0,55	1,15	1,61	9,24	11,07	15,09
6	0,87	1,64	2,20	10,65	12,59	16,81
7	1,24	2,17	2,83	12,02	14,06	18,48
8	1,65	2,73	3,49	13,36	15,51	20,09
9	2,09	3,33	4,17	14,68	16,92	21,67
10	2,56	3,94	4,87	15,99	18,31	23,21
11	3,05	4,58	5,58	17,28	19,68	24,72
12	3,57	5,23	6,30	18,55	21,03	26,22
13	4,11	5,89	7,04	19,81	22,36	27,68
14	4,66	6,57	7,79	21,06	23,69	29,14
15	5,23	7,26	8,55	22,31	25,00	30,58
16	5,81	7,96	9,31	23,54	26,30	32,00
17	6,41	8,97	10,09	24,77	27,59	33,41
18	7,02	9,39	10,86	25,99	28,87	34,81
19	7,63	10,12	11,65	27,20	30,14	36,19
20	8,26	10,85	12,44	28,41	31,41	37,57
21	8,90	11,59	13,24	29,62	32,67	38,93
22	9,54	12,34	14,04	30,81	33,92	40,29
23	10,20	13,09	14,85	32,01	35,17	41,64
24	10,86	13,85	15,66	33,19	36,42	43,98
25	11,52	14,61	16,47	34,38	37,65	44,31
26	12,20	15,37	17,29	35,56	38,89	45,64
27	12,88	16,15	18,11	36,74	40,11	46,96
28	13,56	16,93	18,94	37,92	41,34	48,28
29	14,26	17,71	19,77	39,09	42,56	49,59
30	14,95	18,49	20,60	40,26	43,77	50,89
40	22,16	26,51	29,05	51,81	55,76	63,69
50	29,71	34,76	37,69	63,17	67,51	76,15
100	70,07	77,93	82,36	118,50	124,34	135,81

Додаток 10. Імпорт даних із програми *MS Excel* у програму *MathCad*

I спосіб

1) потрібно скопіювати діапазон клітинок робочого листа програми *MS Excel*, який містить дані для імпортування;

2) скопійовані дані вставляються у програму *MathCad* на місце значення змінної. У результаті цій змінній буде відповідати масив даних.

II спосіб

У програмі *MathCad* існують спеціальні інструменти, які дозволяють працювати з даними *MS Excel* та інших програм. Одним із таких інструментів є функція **READFILE**:

READFILE("file","type",[colwidth,rows,cols,emptyfill]) – повертає матрицю з елементами, які зчитані із зовнішнього файла даних, де "**file**" – назва файла (включаючи шлях до нього на диску); "**type**" – тип файла ("**delimited**" або "**Excel**"); **colwidth** – ширина стовпця даних, який зчитується із файла у випадку вибору попереднього параметра типу "fixed", тобто з фіксованою шириною даних; **rows** – початковий рядок імпорту даних або двокомпонентний вектор, який задає інтервал імпорту рядків; **cols** – початковий стовпець імпорту даних або двокомпонентний вектор, який задає інтервал імпорту стовпців; **emptyfill** – значення, яке буде використано для заміни даних, які відсутні (пустот у файлі). Для нього можна використовувати значення **NaN** (**НеЧисло**). Потрібно відмітити, що ця функція реагує на оновлення початкових даних.

III спосіб

Сюди відносяться декілька інструментів імпортування, які визиваються через пункт меню «**Insert**»:

1) **Insert** → **Data** → **Table**. Створюється таблиця 2х2, щоб імпортувати дані, потрібно натиснути на верхню ліву клітину правою кнопкою миші та вибрати у контекстному меню пункт «**Import**». Далі, за допомогою кнопки «**Browse**» вибрати потрібний файл *MS Excel*, натиснути «**Next**», вибрати потрібний діапазон клітинок і натиснути «**Ok**». Імпортовані дані слід присвоїти змінній;
2) **Insert** → **Data** → **File Input**. Відмінність від попереднього методу полягає у тому, що зміни імпортованих даних у початковому документі *MS Excel* будуть відображатись і при перерахунку у програмі *MathCad*;
3) **Insert** → **Data** → **Data Import Wizard**;
4) **Insert** → **Component**. У вікні «**Component Wizard**» потрібно вибрати Microsoft Excel, потім за допомогою кнопки «**Browse**» вибрати потрібний файл. У розділах «**Inputs**» і «**Outputs**» визначається кількість вхідних та вихідних змінних і діапазон клітинок, якому ці змінні будуть відповідати.