Зміст

Вступ

Задача будь-якої науки полягає у з'ясуванні та дослідженні закономірностей, яким підкоряються реальні явища і процеси. Виявлення та дослідження закономірностей економічних явищ і процесів мають як теоретичне, пізнавальне значення так і широке застосування у плануванні, прогнозуванні та управлінні. Зазвичай на економічні та соціально-економічні явища і процеси впливають різноманітні фактори. У переважній більшості випадків різноманітні закономірності можуть бути виявлені при цілеспрямованому статистичному дослідженні масових явищ і процесів, яке полягає у збиранні даних, їх систематизації і упорядкуванні. Оскільки спостережуваний процес або явище підлягають впливу множини факторів, то їх індивідуальні прояви будуть різнитися. Але у масових сукупностях об'єктів спостережень проявляються загальні закономірності, у формування яких кожна одиниця сукупності вносить свій «вклад».

Теорія імовірностей – це розділ математики, який вивчає закономірності масових випадкових явищ. Її предметом є дослідження взаємозв'язків між випадковими подіями, характеристик величин, числові значення яких змінюються в залежності від випадка, закономірностей поведінки масивів таких випадкових величин. Основним методом теорії імовірностей є побудова ймовірносних моделей, які є частинними випадками математичних моделей. Слід відзначити, що все на світі є випадковим.

Предметом математичної статистики є засоби та прийоми наукового аналізу даних, що відносяться до масових явищ, з метою визначення деяких узагальнюючих ці дані характеристик і виявлення статистичних закономірностей. Основним методом математичної статистики є вибірковий метод.

Розділ 1. Комбінаторика. Події та їх алгебра. Різні означення ймовірності події

Після вивчення даного розділу ви зможете:

- розуміти деякі основні поняття комбінаторики;
- вміти розраховувати кількості різних варіантів прийняття рішень (зокрема, і економічних);
- розуміти, що таке події, їх імовірності та їх смисл;
- виконувати найпростіші розрахунки ймовірностей подій.

Елементи комбінаторики.

Уявимо собі деяку сукупність елементів певної природи – множину. Досить поширеними є задачі складання із елементів скінченних множин комбінацій (або груп) за певними правилами відбору елементів у ці комбінації.

Задачі визначення кількості комбінацій та пошуку алгоритмів для побудови таких комбінацій відносяться до основних задач розділу математики, що називається комбінаторним аналізом або комбінаторикою.

Основними правилами комбінаторики є:

Правило добутку. Нехай потрібно виконати одна за одною деякі k дій. Якщо кожну із цих дій можна виконати $n_i \quad , i = 1,2,...,k$ способами, то усі k дій разом можуть бути виконані $n_1 \cdot n_2 \cdot ... \cdot n_k$ способами.

Для скінченних множин: якщо маємо дві множини $A = \{a_1,...,a_m\}, \quad B = \{b_1,...,b_n\}$. Тоді множина усіх можливих пар $C = \{(a_i, b_j) \mid i = 1,...,m;\ j = 1,...,n\}$ містить $m \cdot n$ елементів.

Правило суми. Якщо дві дії взаємно виключають одна одну, причому одну із них можна виконати m способами, а іншу - n способами. Тоді виконати хоча б одну із цих дій можна $m + n$ способами.

Для скінченних множин: якщо для двох множин A і B (див.вище) $A \cap B = \varnothing$, то множина $A \cup B$ містить $m + n$ елементів.

Приклад: а) із міста А у місто В вантаж можна доставляти 6 шляхами, а з міста В до міста С – 4 шляхами (див. рис.а)). Скількома маршрутами можна доставити вантаж з міста А до міста С?

б) додамо до попередніх умов місто Г (див.рис.б)). Скількома маршрутами можна доставити вантаж з міста А до міста С?

в) банк пропонує наступні кредити: 5 піврічних під 20% річних, 3 річних під 25% річних та 4 півторарічних під 30% річних. Скількома способами можна взяти два різних кредити?

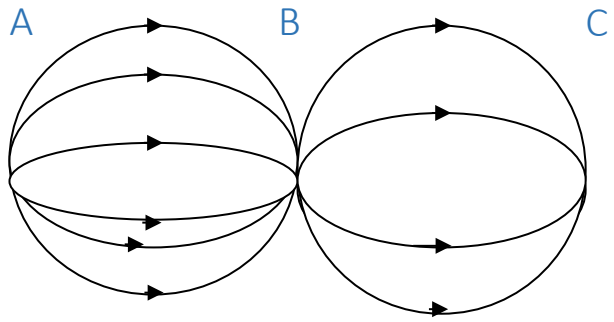


Рис. а)

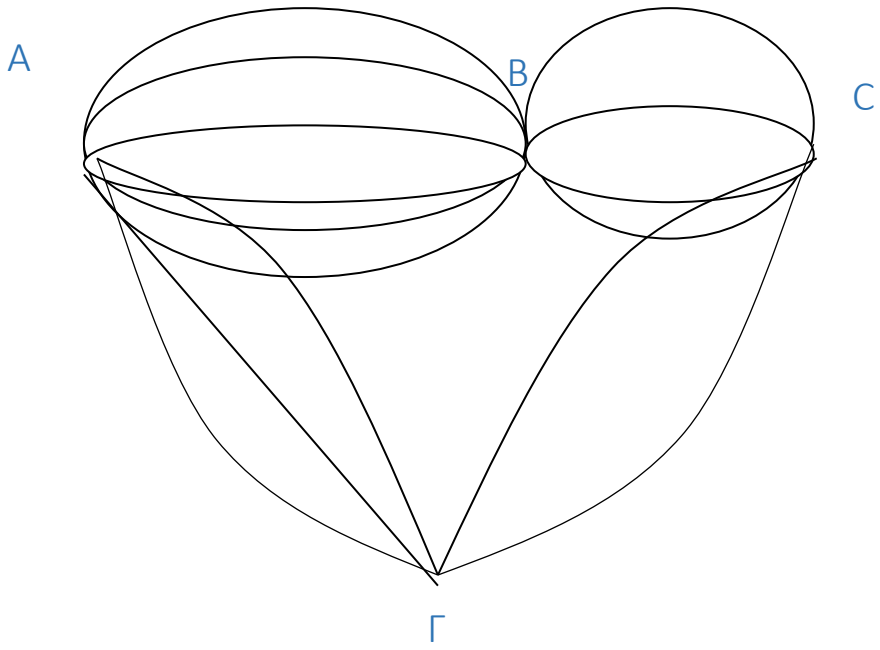


Рис. б)

Розв'язування:

а) вибравши один із 6 шляхів з А до В, далі можемо доставляти вантаж одним із 4 шляхів із В до С. За правилом добутку дістанемо $6 \cdot 4 = 24$ різних маршрутів доставки вантажу із А до С.

б) можливі випадки: маршрут доставки проходить через місто В або через місто Г. Для кожного із цих випадків за правилом добутку відповідна кількість

маршрутів дорівнює 24 та 6. За правилом суми загальна кількість можливих маршрутів доставки вантажу із А до С становить $24+6=30$.

в) можливі три варіанти: перший – взяти піврічний і річний кредити, другий – піврічний і півторарічний, третій – річний та півторарічний. Для кожного із цих варіантів за правилом добутку неважко підрахувати кількості можливостей: 15, 20 та 12. За правилом суми остаточно дістаємо загальну кількість способів: 15+20+12=47.

Переставлення (перестановки).

Нехай потрібно підрахувати число способів, за якими можна розмістити в ряд n елементів множини, тобто кожне розміщення є скінченною множиною, елементи якої записано у певному порядку.

Скінченні множини, для яких істотним є порядок елементів, називають ***впорядкованими***. Вказати порядок розміщення елементів у скінченій множині означає нумерацію елементів множини. Дві впорядковані множини рівні, якщо вони складаються із однакових елементів і однаково впорядковані. Наприклад, множини (а,в) і (в,а) – різні впорядковані множини, елементами яких є елементи невпорядкованої множини {а,в}.

Означення. Будь-яка впорядкована множина, що складається із n елементів називається ***перестановкою із n елементів***.

Перестановки складаються з одних і тих самих елементів, а відрізняються лише порядком елементів.

Число перестановок у множині із n елементів позначається P_n і обчислюється за формулою: $P_n = n! = n\cdot(n-1)\cdot\ldots\cdot 2\cdot 1$.

Для числа перестановок справедлива наступна рекурентна формула:

$$P_n = n\cdot P_{n-1}.$$

Зауважимо, що число перестановок можна підраховувати в Excel за допомогою функції «ФАКТР».

Приклад. Скільки різних трьохзначних чисел можна скласти, використовуючи у числах цифри 1,2,3 не більше одного разу? Вкажіть ці числа.

Розв'язування. Викликавши в Excel функцію «ФАКТР» з аргументом 3, отримуємо шукану кількість чисел. Дійсно, це числа 123, 132, 213, 231, 312, 321.

Розміщення.

Нехай маємо множину із n різних елементів.

Означення. **Розміщенням із n елементів по k** називають підмножини, що складаються із k елементів, вибраних із даних n елементів і розміщених у певному порядку (іншими словами – всі впорядковані підмножини даної множини).

Розміщення можуть відрізнятись одне від одного або самими елементами, або їх порядком.

Кількість розміщень із n елементів по k позначають $A_n^k \quad , k \le n$ і обчислюють за формулою: $A_n^k = \frac{n!}{(n-k)!} = n \cdot (n-1) \cdot \ldots \cdot (n-k+1)$.

Для кількості розміщень справедлива наступна рекурентна формула:

$$A_n^{k+1} = A_n^k \cdot (n-k) \quad , k \le n-1.$$

Зауважимо, що число розміщень можна підраховувати в Excel за допомогою функції «ПЕРЕСТР».

Приклад. Скільки різних чисел можна скласти, використовуючи у числах цифри 1,2,3 не більше одного разу? Вкажіть ці числа.

Розв'язування. Кількість (6) різних тризначних чисел (яку можна знайти за допомогою функції «ПЕРЕСТР» з аргументами 3 і 3) була знайдена вище. Кількість різних двозначних чисел знайдемо за допомогою функції «ПЕРЕСТР» з аргументами 3 і 2. Вона дорівнює 6, це числа: 12, 13, 21, 23, 31, 32. Кількість різних однозначних чисел знаходимо за допомогою функції «ПЕРЕСТР» з аргументами 3 і 1, і вона дорівнює 3. Це 1, 2, 3. Отже кількість різних чисел за умовою задачі становить 15.

Сполучення (комбінації).

Означення. Будь-яка підмножина із m елементів даної множини, яка містить n елементів, називається **комбінацією із n елементів по m**.

Комбінації різняться складом елементів.

Число комбінацій позначають C_n^m і обчислюють за формулою:

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

Зауважимо, що число комбінацій можна підраховувати в Excel за допомогою функції «ЧИСЛКОМБ».

Для кількості комбінацій справедливі наступні формули:

а) $C_n^0 + C_n^1 + ... + C_n^n = 2^n$ (наслідок біноміальної формули Ньютона).

б) $C_n^m = C_n^{n-m}$ (властивість симетрії).

в) $C_n^m + C_n^{m+1} = C_{n+1}^{m+1}$, де $0 \le m < n$ (рекурентне співвідношення або правило Паскаля).

Між кількостями розміщень, переставлень та комбінацій існує очевидний зв'язок: $A_n^m = P_m \cdot C_n^m$.

Приклад. Скількома способами можна вибрати трьох делегатів на конференцію з групи із 10 студентів?

Розв'язування. Кількість способів обрання 3 делегатів із 10 підрахуємо в Excel за допомогою функції «ЧИСЛКОМБ» із аргументами 10 і 3. Таких способів 120.

Події. Класифікація подій.

Під **випробуванням (дослідом, експериментом)** розуміють відтворення (реалізацію) певного комплекса умов, які можна повторювати необмежену кількість разів.

Під **подією** розуміють можливий наслідок (результат) випробування. Події, як правило, позначають великими літерами.

Означення. **Випадковою подією** називається подія, яка може настати (з'явитись) або не настати у даному випробуванні. Всюди надалі для скорочення слово випадкова опускатимемо.

Означення. **Достовірною подією** називається подія, яка обов'язково настає у даному випробуванні. Достовірну подію позначатимемо U.

Означення. **Неможливою подією** називається подія, яка не може настати у даному випробуванні. Неможливу подію позначатимемо V.

Означення. **Попарно несумісними подіями (несумісними у сукупності)** називаються події $A_1, A_2, ..., A_n$, якщо у даному випробуванні ніякі дві з них не можуть настати разом (поява однієї із подій виключає появу будь-якої іншої). У супротивному випадку події називають **сумісними**.

Означення. **Єдино можливими подіями** називаються події $A_1, A_2, ..., A_n$, якщо у даному випробуванні обов'язково настане хоча б одна із цих подій.

Означення. Події $A_1, A_2, ..., A_n$ утворюють **повну групу**, якщо вони єдино можливі та попарно несумісні.

Означення. Дві події A і $\bar{A}$, які утворюють повну групу, називаються взаємно **протилежними**.

Означення. Події $A_1, A_2, ..., A_n$ називаються **рівноможливими**, якщо вони мають однакові шанси до появи у даному випробуванні.

Означення. Кажуть, що подія A сприяє події B, якщо у даному випробуванні в результаті появи події A обов'язково з'явиться (настане) подія B.

Означення. **Простір елементарних подій - – усі єдино можливі, рівноможливі та несумісні події, які неможливо поділити на більш прості події**.

Приклади.

Алгебра подій.

Означення. Добутком (перетином) двох подій A і B (позначається $A \cdot B$ або $A \cap B$) називається подія, яка полягає у одночасній появі подій A і B у даному випробуванні. Означення легко розповсюджується на випадок скінченної кількості співмножників – подій.

Означення. Сумою (об'єднанням) двох подій A і B (позначається $A + B$ або $A \cup B$) називається подія, яка полягає у появі хоча б однієї із цих подій (події A або події B, або одночасно A і B разом) у даному випробуванні. Означення нелегко розповсюджується на випадок скінченної кількості доданків – подій.

Означення. Різницею двох подій A і B (позначається $A-B$ або $A\setminus B$) називається подія, яка полягає в тому, що настане подія A, а подія B не настане у даному випробуванні.

Для наглядності при зображенні різноманітних подій та дій над ними користуються так званими діаграмами Венна-Ейлера. При цьому прямокутник зображає так звану **універсальну подію - простір елементарних подій**.

Приклади.

Класичне означення імовірності.

Означення. Імовірність події A дорівнює:

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

де n - число (кількість) подій у просторі елементарних подій,

а m - число наслідків (із простору елементарних подій), які сприяють появі події A.

Приклади.

Властивості:

1. Для довільної події A: $0 \le P(A) \le 1$;
2. Для достовірної події U: $P(U)=1$.
3. Для неможливої події V: $P(V)=0$.

Приклади.

Геометричне означення імовірності.

Означення. Імовірність події A дорівнює:

$$P(A)=\frac{mes(A)}{mes(U)},$$

де $mes(U)$ - геометрична міра простору елементарних подій,

а $mes(A)$ - геометрична міра частини простору елементарних подій, яка сприяє появі події A.

Приклади.

Статистичне означення імовірності.

Означення. Нехай проводиться n випробувань (які можна повторювати при незмінних умовах необмежено). **Частотою** $m(A)$ називається кількість випробувань (із n) , в яких з'явилась подія A. **Відносною частотою (або часткою)** $w(A) = \frac{m(A)}{n}$ називається відношення частоти появи події A до загальної кількості випробувань.

Означення. Статистичною імовірністю події A називають число, що характеризує можливості появи події і яке дорівнює:

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

Приклади.

Розділ 2. Умовна імовірність. Теореми добутку, додавання та наслідки з них. Формули повної імовірності та Байєса

Після вивчення даного розділу ви зможете:

- розуміти, що таке незалежні та залежні події, обчислювати їх імовірності;
- вміти розкладати складні події у комбінації простих;
- застосовувати теореми додавання та добутку, формули повної ймовірності та Байєса;
- виконувати ймовірнісне моделювання економічних та практичних задач.

Означення. Події A і B називаються **незалежними**, якщо поява або непоява однієї з них не впливає на імовірність настання іншої. У супротивному випадку події називаються **залежними**.

Приклад. В урні знаходяться 3 білих і 3 чорних кулі. Із урни навмання дістають одну кулю і розглядають подію A - куля біла. Потім цю кулю повертають до урни (схема «повернених куль») і дістають навмання одну із куль. Розглядається подія B - куля чорна. Очевидно, що у цьому випадку імовірність появи події B : $P(B)=\frac{3}{6}$ і не залежить від появи чи непояви події A, оскільки події незалежні.

Приклад. В урні знаходяться 3 білих і 3 чорних кулі. Із урни навмання дістають одну кулю і розглядають подію A - куля біла. Першу витягнуту кулю не повертають до урни (схема «неповернених куль») і дістають навмання наступну кулю. Розглядається подія B - куля чорна. Очевидно, що у цьому випадку імовірність появи події B буде залежати від того, якого кольору була перша куля, тобто необхідно розглядати умовні імовірності:

а) імовірність події B при умові, що настала подія A: $P_A(B)=P(B/A)=\frac{3}{5}$;

б) імовірність події B при умові, що не настала подія A (тобто, настала подія $\bar{A}$): $P_{\bar{A}}(B)=P(B/\bar{A})=\frac{2}{5}$;

Отже у цьому прикладі події залежні.

Означення. Події $A_1, A_2, ..., A_n$ називаються **незалежними у сукупності** (або просто **незалежними**), якщо імовірність появи кожної з них не залежить від того, відбулись чи ні будь-які інші події. У супротивному випадку події $A_1, A_2, ..., A_n$ називаються **залежними**.

Теорема добутку.

Теорема (добутку імовірностей). Імовірність добутку двох подій A і B дорівнює добутку імовірності однієї з них на умовну імовірність іншої, при умові, що настала перша подія:

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

Доведення. Нехай n - кількість подій (елементарних наслідків) у просторі елементарних подій, з яких k подій сприяють появі AB, m - сприяють появі A, а l - сприяють появі B. За класичним означенням імовірності:

$$P(AB) = \frac{k}{n} = \frac{k \cdot m}{n \cdot m} = \frac{m}{n} \cdot \frac{k}{m} = P(A) \cdot P_A(B).$$

Аналогічно доводиться, що $P(AB) = P(B) \cdot P_B(A)$.

Теорема легко розповсюджується на випадок фіксованої кількості співмножників-подій. Наприклад, для трьох подій:

$$P(ABC) = P(A) \cdot P_A(B) \cdot P_{AB}(C).$$

Наслідок 1 (формули визначення умовних імовірностей). Якщо імовірності подій відмінні від нуля, то

$$P_A(B) = P(AB) / P(A) \quad , \quad P_B(A) = P(AB) / P(B).$$

Зауважимо, що теорема добутку справедлива навіть у випадку нульових імовірностей подій.

Наслідок 2. Якщо подія A не залежить від події B, то і навпаки, подія B не залежить від події A, тобто вони взаємно незалежні.

Наслідок 3. Із незалежності подій A і B випливає незалежність пар подій : A і $\bar{B}$, $\bar{A}$ і B, $\bar{A}$ і $\bar{B}$.

Наслідок 4. Імовірність добутку двох незалежних подій дорівнює добутку їх імовірностей:

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

Наслідок легко розповсюджується на випадок фіксованої кількості співмножників-подій.

Приклад .

Теорема додавання.

Теорема. Імовірність суми двох подій A і B дорівнює сумі імовірностей цих подій без імовірності їх добутку. Іншими словами, імовірність появи хоча б однієї із двох подій дорівнює сумі їх імовірностей без імовірності їх сумісної появи:

$$P(A+B)=P(A)+P(B)-P(AB).$$

Доведення. За класичним означенням :

$$P(A+B)=\frac{m+l-k}{n}=\frac{m}{n}+\frac{l}{n}-\frac{k}{n}=P(A)+P(B)-P(AB).$$

Зауважимо, що теорема досить важко розповсюджується на випадок скінченної кількості доданків-подій. Так, наприклад, для трьох подій:

$$P(A+B+C)=P(A)+P(B+C)-P(A(B+C))=$$

$$=P(A)+P(B)+P(C)-P(AB)-P(AC)-P(BC)+P(ABC)$$

.

Наслідок 1. Імовірність суми двох несумісних подій дорівнює сумі їх імовірностей:

$$P(A+B)=P(A)+P(B).$$

Наслідок легко розповсюджується на випадок фіксованої кількості несумісних подій-доданків.

Наслідок 2. Сума імовірностей подій $A_1, A_2,..., A_n$, що утворюють повну групу, дорівнює одиниці:

$$P(A_1)+P(A_2)+...+P(A_n)=1.$$

Наслідок 3. Для взаємно протилежних подій A і $\bar{A}$:

$$P(A)+P(\bar{A})=1,\quad P(A)=1-P(\bar{A}),\quad P(\bar{A})=1-P(A).$$

Наслідок 4 . Імовірність появи хоча б однієї із подій $A_1, A_2,..., A_n$ дорівнює:

$$P(A_1 + A_2 + ... + A_n) = 1 - P(\bar{A}_1 \cdot \bar{A}_2 \cdot ... \cdot \bar{A}_n).$$

Зокрема, якщо події незалежні в сукупності, то:

$$P(A_1 + A_2 + ... + A_n) = 1 - P(\bar{A}_1) \cdot P(\bar{A}_2) \cdot ... \cdot P(\bar{A}_n).$$

Дерево імовірностей. Дерево рішень.

Дерево імовірностей.

Обчислювати імовірності складних подій за класичним означенням буває досить складно, а іноді взагалі неможливо. Тому доводиться використовувати теореми додавання та добутку. При цьому важливо ураховувати всі можливі наслідки, для чого складається так зване «дерево імовірностей», на якому випробування позначають кругами, а можливі наслідки-події – лініями-«гілками».

Приклад. При прийомі хворих в лікарні встановлено, що 80% пацієнтів відправляють додому після обстеження та надання необхідної допомоги. Інші 20% розміщують наступним чином: 60% попадають до корпусу А і 40% - до корпусу В. Щоденно лікар Синиця оглядає 70% пацієнтів корпусу А і тільки 10% пацієнтів корпусу В. Лікар Руденко оглядає усіх інших пацієнтів. Які імовірності того, що пацієнт, який надійшов до лікарні, буде під наглядом того чи іншого лікаря?

Розв'язування. Зобразимо дану ситуацію, скористувавшись деревом імовірностей:

Дерево рішень.

Дерево рішень – це графічне зображення ситуації, яка має декілька альтернативних рішень. Його складовими частинами є «рішення» (зображаються прямокутниками) та «ймовірністні події» (круги). Дерево рішень дозволяє уявити конкретну проблему та встановити імовірності настання подій та їх очікуванних значень. Такі діаграми призводять до побудови більш простих дерев імовірностей, що пов'язані з послідовностями наслідків. Дерево рішень ілюструє результати з точки зору критичних факторів (таких, як прогнозні доходи та витрати тощо).

Приклад. Дехто володіє акціями вартістю 1000 у.о. Він повинен прийняти рішення відносно того, чи тримати йому акції, або продати їх усі, або придбати ще акції на суму 500 у.о. Ймовірність 20% росту курсової вартості акцій становить 0,6 , а ймовірність зниження курсової вартості на 20%

- 0,4. Яке рішення необхідно прийняти, щоб максимізувати очікуваний прибуток?

Розв'язування. ОПР (особі, що приймає рішення) потрібно обрати один із трьох варіантів: продати усі акції, тримати їх або купити ще. Зобразимо ці варіанти за допомогою діаграми - дерева рішень і підрахуємо для кожного із варіантів очікувані значення прибутків:

Формула повної ймовірності.

Теорема. Нехай подія A може настати лише сумісно з хоча б однією із подій-гіпотез $H_1, H_2, ..., H_n$, які утворюють повну групу подій. Тоді імовірність (повна імовірність) події A дорівнює:

$$P(A) = \sum_{i=1}^{n} P(H_i) P_{H_i}(A) = P(H_1) P_{H_1}(A) + ... + P(H_n) P_{H_n}(A)$$

,

тобто сумі добутків імовірностей гіпотез на умовні ймовірності події, при умові, що настала відповідна гіпотеза.

Доведення. Оскільки події H_i утворюють повну групу (є незалежними, несумісними та єдиноможливими), то $A = A \cdot (H_1 + H_2 + \ldots + H_n) = A \cdot H_1 + A \cdot H_2 + ... + A \cdot H_n$. Застосовуючи теореми суми та добутку, отримаємо:

$$P(A) = P(A \cdot H_1 + A \cdot H_2 + ... + A \cdot H_n) = P(A \cdot H_1) + P(A \cdot H_2) + ... + P(A \cdot H_n) =$$

$$= P(H_1) P_{H_1}(A) + P(H_2) P_{H_2}(A) + ... + P(H_n) P_{H_n}(A)$$

Приклад . Кожна із двох урн містить по 3 білих та 5 чорних куль. Із першої урни до другої переклали дві кулі. Знайти імовірність того, що навмання взята із другої урни куля буде біла.

Розв'язування.

Приклад. Статистика запитів на отримання кредитів у банку наступна: 20% - державні органи, 30% - інші банки, інші – фізичні особи. Дослідженнями встановлено, що імовірності неповернення кредитів відповідно дорівнюють 0,01 ; 0,05 ; 0,2 . Знайти імовірність неповернення чергового кредита.

Розв'язування.

Формули Байєса.

Теорема. Нехай подія A може настати лише сумісно з хоча б однією із подій-гіпотез $H_1, H_2, ..., H_n$, які утворюють повну групу. Якщо подія A настала, то умовні (уточнені) імовірності гіпотез дорівнюють:

$$P_A(H_i) = \frac{P(H_i)P_{H_i}(A)}{P(A)}, i = 1,2,...,n;$$

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

Доведення. Зазначимо, що виконуються усі умови теореми – формули повної ймовірності. Розглянемо одну із подій AH_i і скористуємось теоремою добутку:

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

Звідси:

$$P_A(H_i) = P(H_i)P_{H_i}(A) / P(A),$$

де $P(A)$ - повна імовірність.

Зауваження. Доведені формули називають формулами переоцінки гіпотез. В них приймають участь апріорні (до випробування) імовірності $P(H_i)$ гіпотез та їх апостеріорні (після випробування) ймовірності $P_A(H_i)$, тобто формули дають можливість переоцінити ймовірності гіпотез після настання події.

Приклад. За умовами попереднього прикладу до банку надійшло повідомлення по факсу про неповернення чергового кредиту, але у повідомленні атрибути клієнта погано відпечатались. Визначити, до якої категорії клієнтів (державні органи, інші банки, фізичні особи) імовірніше за все належить "неповерненик".

Розв'язування.

Розділ 3. Дискретні випадкові величини (ДВВ). Закон розподілу ДВВ. Операції над незалежними ДВВ. Числові характеристики ДВВ

Поняття випадкової величини (ВВ) – є одним із фундаментальних в теорії ймовірностей. ВВ на відміну від події (яка характеризує результат випробування), є кількісною характеристикою випадкового результату випробування.

Означення. Випадковою величиною (ВВ) називається величина, яка в результаті випробування в залежності від випадкових обставин може набувати деякого (але тільки одного) значення.

Приклад. а) число попадань в мішень із 10 пострілів – ВВ, яка може набувати значень 0,1,2,…,10;

б) число пострілів у мішень до першого попадання – ВВ, яка може приймати значення 1,2,3,… ;

в) відстань від центра круглої (радіуса R) мішені до точки попадання в неї – ВВ, яка може приймати будь-яке (але тільки одне) значення з проміжка [0;R].

ВВ позначають великими літерами, а їх значення – відповідними малими з певними індексами. Приклад показує, що є дискретні (а) і б)) та неперервні (в)) ВВ.

Означення. Дискретною випадковою величиною (ДВВ) називається ВВ, яка може приймати окремі ізольовані значення з певними ймовірностями (причому кількість можливих значень або скінченна, або нескінченна, але злічена). На відміну від ДВВ, значення неперервних ВВ повністю заповнюють деякий проміжок (скінченний або нескінченний).

ВВ вважається заданою, якщо задано її закон розподілу.

Означення. Законом розподілу ДВВ називається відповідність між множиною її можливих значень та відповідними імовірностями (тобто, імовірностями, з якими можуть набуватись ці значення).

Основними способами задання законів розподілу ДВВ є табличний, графічний та аналітичний.

ДВВ задано таблицею:

X	x_1	x_2	. . .	x_n
P	p_1	p_2	. . .	p_n

де x_i, $i = 1,2,...,n$ – можливі (різні) значення ВВ X, а $p_i = P(X = x_i)$ – відповідні ймовірності, причому, $p_1 + p_2 + \ldots + p_n = 1$, оскільки події $X = x_i$ утворюють повну групу. Останню умову

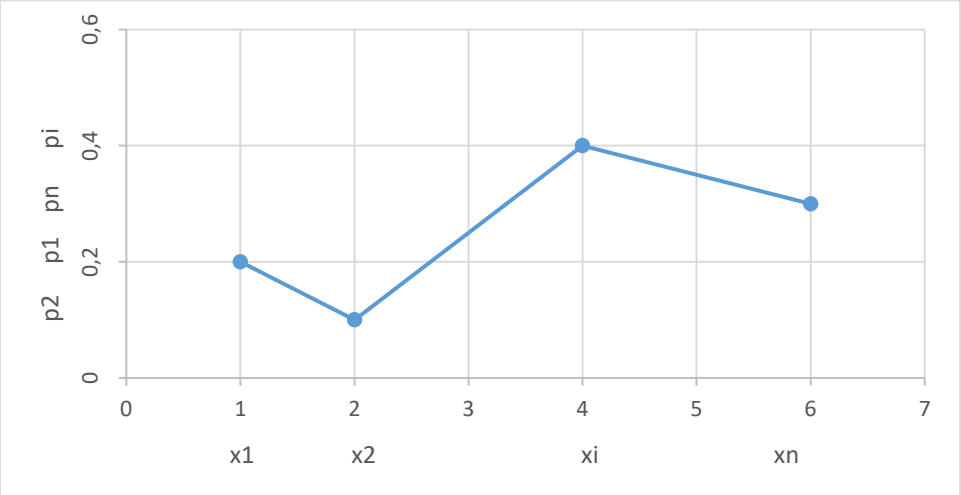
часто називають **основною властивістю розподілу** або **умовою нормування** ДВВ.

Зауважимо, що у випадку зліченої множини значень ДВВ:

X	x_1	x_2	…	x_n	…
P	p_1	p_2	…	p_n	…

умовою нормування буде збіжність до одиниці ряду відповідних імовірностей: $\sum_{i=1}^{\infty} p_i = 1$.

Наочною формою задання ДВВ є графічний спосіб, при якому в системі координат XoP відкладають точки $(x_i; p_i)$ і з'єднують їх відрізками:



Отриману фігуру називають **полігоном** розподілу ймовірностей або многокутником розподілу. Зауважимо, що на проміжках $(x_i; x_{i+1})$ ВВ X не набуває значень, тому імовірності появи її можливих значень дорівнюють нулю, а з'єднання точок відрізками робиться для наочності.

При аналітичному способі задання закону розподілу ДВВ X вказують формулу (функцію), за якою знаходяться відповідні ймовірності $p_i = P(X = x_i) = f(x_i)$, або задають так звані функції розподілу.

Приклад . Імовірності попадання в мішень для першого стрілка – 0,8 , а для другого – 0,9. Обидва стрілки роблять по одному пострілу. Скласти закон розподілу ВВ – кількості попадань в мішень, побудувати полігон розподілу.

Розв'язування.

ОПЕРАЦІЇ НАД ДВВ.

Дві ВВ називаються **незалежними**, якщо закон розподілу однієї із них не залежить від того, які можливі значення прийняла інша ВВ.

Нехай незалежні ДВВ X та Y задані таблицями розподілу:

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

Y	y_1	y_2	...	y_m
P	q_1	q_2	...	q_m

Добутком ВВ X на сталий множник k називається ВВ kX , яка набуває можливих значень kx_i з тими ж імовірностями p_i, що і ВВ X.

k-им степенем ($k = 2,3,...$) **ВВ X** називається ВВ X^k, яка приймає значення x_i^k з тими ж імовірностями p_i, що і ВВ X.

Сумою (різницею або добутком) двох незалежних ВВ X та Y називається ВВ $X+Y$ ($X-Y$ або XY), яка приймає всі можливі значення $x_i + y_j$ ($x_i - y_j$ або $x_i y_j$) з імовірностями p_{ij}, що знаходяться за теоремою добутку: $p_{ij} = p_i q_j = P(X = x_i)P(Y = y_j)$.

Приклад. ДВВ задані таблицями розподілу

X	0	1	2
P	0,1	0,2	0,7

Y	1	3
P	0,6	0,4

Знайти закони розподілу ВВ $X-Y,\ XY,\ 2X,\ X+X,\ XX,\ X^2$.

Розв'язування.

ЧИСЛОВІ ХАРАКТЕРИСТИКИ ДВВ.

У багатьох практичних задачах немає необхідності мати закон розподілу ВВ, а достатньо знати лише деякі її **числові характеристики: математичне сподівання, дисперсію, середнє квадратичне відхилення, початкові та центральні моменти**.

Нехай ДВВ X задана таблицею:

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

Означення. Математичним сподіванням (середнім значенням або центром розподілу) ДВВ X називається сума добутків всіх її значень на відповідні ймовірності, тобто

$$M(X)=E(X)=\overline{X}=a=\sum_{i=1}^{n}x_i p_i .$$

Зауваження. Для будь-якої ДВВ із скінченною множиною значень математичне сподівання існує (і є **невипадковим, сталим числом**) і має таку саму розмірність, що й сама ДВВ. У випадку нескінченної зліченої множини значень ДВВ її математичне сподівання (МС) визначається як сума ряда, який може розбігатись, і тому МС може не існувати.

Приклад. Згідно статистичним даним, імовірність смерті 25-річної людини протягом року дорівнює 0,008. Страхова компанія пропонує застрахувати життя на 5000грн. Якою повинна бути величина річного внеску, щоб ця страховка була для компанії незбитковою?

Розв'язування.

Приклад. (дохідність портфеля цінних паперів). Дохідність портфеля характеризується середньозваженою дохідністю його складових. Наприклад, середня очікувана дохідність $M(R)$ портфеля із двох активів A і B розраховується як середньо-зважена його складових:

$$M(R)=w_A M(R_A)+w_B M(R_B),$$

де w_A, w_B - питомі вагові коефіцієнти активів ($w_A \geq 0, w_B \geq 0, w_A + w_B = 1$), а $M(R_A), M(R_B)$ - їх середні очікувані дохідності.

Приклад (середня дохідність фінансової операції). Перший варіант фінансової операції передбачає початкові витрати – інвестиції розміром 10000грн та отримання прибутку в 3000грн з імовірністю 0,9. Другий варіант при витратах 20000грн дає прибуток в 10000грн з імовірністю 0,1. Якою буде середня очікувана дохідність всієї фінансової операції?

Розв'язування. У детермінованому фінансовому аналізі дохідність операції визначається як $d = \frac{K - H}{H}$, де H - грошова оцінка початку операції (початкові витрати, інвестиції) , K - грошова оцінка кінця операції (дохід, нарощений капітал), а $\Delta = K - H$ - прибуток від операції.

ВЛАСТИВОСТІ МАТЕМАТИЧНОГО СПОДІВАННЯ.

1. МС сталої ВВ дорівнює самій цій сталій: $M(c) = c$.

2. Сталий множник виноситься за знак МС: $M(cX) = cM(X)$.

3. МС суми ВВ дорівнює сумі їх МС: $M(X + Y) = M(X) + M(Y)$.

Доведення. Для спрощення розглянемо ДВВ, задані таблицями:

X	x_1	x_2
P	p_1	p_2

Y	y_1	y_2
P	q_1	q_2

Наслідок. МС різниці ВВ дорівнює різниці їх МС: $M(X - Y) = M(X) - M(Y)$.

4. МС добутку (незалежних) ВВ дорівнює добутку їх МС: $M(XY) = M(X)M(Y)$.

Властивості **3,4** та наслідок легко розповсюдити на випадок фіксованої кількості доданків (співмножників), зокрема, неважко довести, що МС середнього арифметичного ВВ дорівнює середньому арифметичному їх МС.

5. МС **центрованої ВВ** $X - M(X)$ дорівнює нулю: $M[X - M(X)] = 0$.

ДИСПЕРСІЯ ДВВ ТА ЇЇ ВЛАСТИВОСТІ. СЕРЕДНЄ КВАДРАТИЧНЕ ВІДХИЛЕННЯ.

Означення. Дисперсією ДВВ називається МС квадрата відхилення ВВ від свого МС, тобто:

$$D(X) = \sigma^2(X) = \mathrm{var}(X) = M[X - M(X)]^2.$$

Дисперсія (якщо вона існує) має розмірність квадрата ВВ, є **невипадковою сталою невід'ємною величиною**, що характеризує розсіювання значень ВВ від центру розподілу – МС.

Для того, щоб мати аналогічну характеристику такої ж розмірності як сама ВВ, розглядають **середнє квадратичне відхилення (стандартне відхилення або стандарт):** $\sigma(X) = \sqrt{D(X)}$.

ВЛАСТИВОСТІ ДИСПЕРСІЇ.

1.(Формула знаходження дисперсії) Дисперсію можна знаходити за формулою: $D(X) = M(X^2) - M^2(X)$.

2. Дисперсія сталої дорівнює нулю: $D(c) = 0$.

3. Сталий множник виноситься за знак дисперсії в квадраті: $D(cX) = c^2 D(X)$.

4. Дисперсія суми незалежних ВВ дорівнює сумі їх дисперсій: $D(X + Y) = D(X) + D(Y)$.

Наслідок. Дисперсія різниці незалежних ВВ дорівнює сумі їх дисперсій: $D(X-Y)=D(X)+D(Y)$.

Наслідок. **Дисперсія центрованої ВВ** $X-M(X)$ співпадає із дисперсією самої ВВ X, а **дисперсія стандартизованої ВВ** $Z=\dfrac{X-M(X)}{\sigma}$ дорівнює одиниці.

5. Якщо ВВ X та Y залежні, то: $D(X\pm Y)=D(X)+D(Y)\pm 2\operatorname{cov}(X,Y)$, де $\operatorname{cov}(X,Y)=M[(X-M(X))(Y-M(Y))]=M(XY)-M(X)M(Y)$ - **коваріація** між ВВ X та Y.

6. Дисперсія добутку незалежних ВВ дорівнює: $D(XY)=M(X^2)M(Y^2)-M^2(X)M^2(Y)$.

7. Дисперсія середнього арифметичного незалежних ВВ дорівнює:

$$D\left(\frac{X_1+\ldots+X_n}{n}\right)=\frac{D(X_1)+\ldots+D(X_n)}{n^2}.$$

Важливий висновок. При збільшенні кількості ВВ їх сумісна дисперсія лише збільшується. Єдиний спосіб зменшити розсіювання – це використовувати середні величини!!! Доречі, варіація (або розсіювання) часто використовується у якості кількісної оцінки ризиків.

Приклад. Протягом деякого року середньомісячна зарплата X (тис.грн) приватного підприємства змінювалась за законом:

X	4,5	4,6	4,7	4,8	4,9	5	5,1	5,2	5,3	5,4	5,5	5,6
P	0,09	0,01	0,08	0,2	0,18	0,06	0,08	0,06	0,09	0,01	0,05	0,09

Обчислити числові характеристики випадкової величини X (середньорічну середньомісячну зарплату та її розсіювання протягом року).

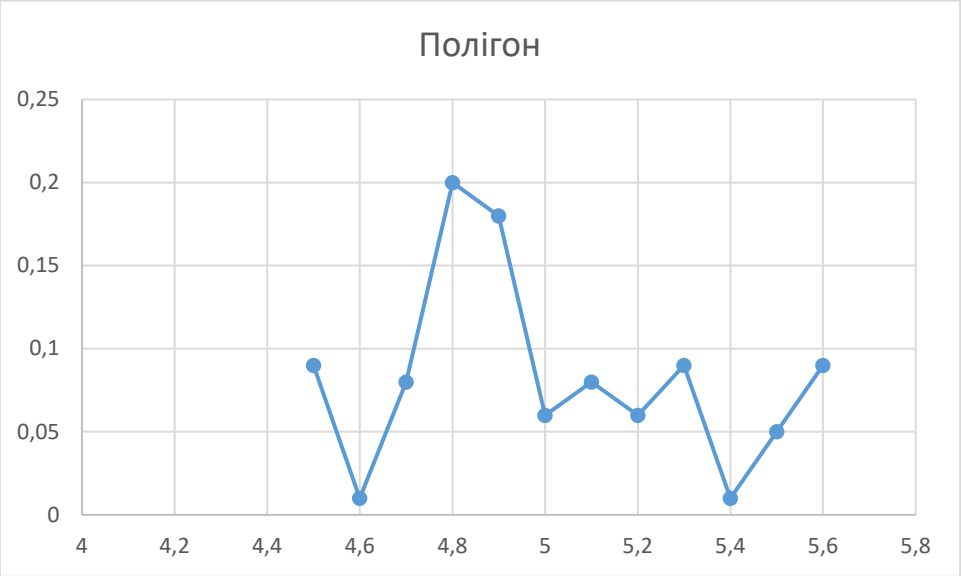
Розв'язування. Скористуємось функціями Excel «СУММПРОИЗ» або «Автосумма» (при роботі з електронними таблицями дані зручніше задавати у стовпцях):

M(X)= «СУММПРОИЗ»($x_i p_i$)=4,999 (тис.грн) - середньорічна середньомісячна зарплата;

D(X)=M(X^2)-M^2(X)= «СУММПРОИЗ»($x_i{}^2 p_i$)-$4{,}999^2$=0,098899 (тис.грн^2)???

Стандартне відхилення становить 0,314482 (тис.грн) або приблизно 315 грн.

Побудуємо полігон розподілу середньомісячної зарплати:



Як видно із полігону, середньорічна середньомісячна зарплата знаходиться у районі тих значень, які мають найбільшу вагу (ймовірність) у розподілі.

МОМЕНТИ РОЗПОДІЛУ ТА ПОВ'ЯЗАНІ З НИМИ ЧИСЛОВІ ХАРАКТЕРИСТИКИ ВВ.

Означення. **Початковим моментом порядку k ВВ X** називають МС ВВ X^k і позначають

$$\nu_k = M(X^k), \quad k = 1,2,...,n.$$

Означення. **Центральним моментом порядку k ВВ X** називають МС ВВ $\left[X - M(X)\right]^k$ і позначають

$$\mu_k = M\left(\left[X - M(X)\right]^k\right), \quad k = 1,2,...,n.$$

Моменти мають властивість: якщо існує момент порядку k ВВ X, то існують її моменти усіх порядків $m < k$.

Відмітимо, що $\nu_1 = M(X), \quad \nu_2 = M(X^2)$;

$$\mu_1 = M\left(\left[X - M(X)\right]\right) = 0,$$

$$\mu_2 = M\left(\left[X - M(X)\right]^2\right) = D(X) = \nu_2 - \nu_1^2.$$

Початковий момент першого порядку дорівнює математичному сподіванню ВВ X і характеризує «центр розподілу». Центральні моменти використовують для характеристики розсіювання значень ВВ відносно її центра розподілу – МС (напр., дисперсія $\mu_2 = D(X)$).

Центральний момент третього порядку μ_3 застосовують для оцінювання асиметрії (скісності) закону розподілу відносно прямої, що паралельна осі ординат і проходить через МС. Для цього використовують безрозмірну величину – **коефіцієнт асиметрії:**

$$As = \frac{\mu_3}{\sigma^3}.$$

Якщо розподіл ВВ симетричний відносно МС, то його $As = 0$. Якщо $As > 0$, то у розподілі «довга частина» полігона розташована праворуч МС – скошеність вправо, а у випадку $As < 0$ спостерігається скошеність розподілу вліво.

Центральний момент четвертого порядку μ_4 використовують для оцінювання крутості (гостровершинності або плосковершинності) розподілу за допомогою **коефіцієнта ексцесу:**

$$Es = \frac{\mu_4}{\sigma^4} - 3.$$

Число 3 віднімається від частки, оскільки для нормального розподілу (розглядатиметься далі), який зустрічається найчастіше, $\mu_4 = 3\sigma^4$ тому $Es = 0$.

Зазначимо, що використовуються також **абсолютні початкові та центральні моменти порядку k ВВ X** , які визначаються як МС випадкових величин $|X|^k$ та $|X - M(X)|^k$.

Модою ДВВ X називається таке можливе її значення $Mo = x_m$, якому відповідає найбільша імовірність, тобто: $p_m = P(X = x_m) = \max_i P(X = x_i)$ (для неперервних ВВ мода визначається як точка локального максимуму щільності розподілу $f(x)$). Якщо розподіл імовірностей (або щільність) має один максимум, то він

називається **унімодальним**. Бувають також бімодальні та мультимодальні розподіли, а також такі, що не мають моди – антимодальні. Для унімодального розподілу мода є, у певному розумінні найімовірнішим значенням.

Медіаною ВВ X називається таке значення $Me(X)$, яке «ділить» розподіл навпіл, тобто:

$$P\big(X < Me(X)\big) = P\big(X > Me(X)\big) = 0{,}5.$$

Відмітимо, що математичне сподівання ВВ може не існувати, а медіана існує завжди і має властивість: $\sum_i \left|x_i - Me(X)\right| = min$, тобто сума абсолютних величин відхилень значень ДВВ від медіани менша, ніж від будь-якої іншої величини. Ця властивість медіани часто використовується на практиці.

Для описування ВВ застосовуються також інші її числові характеристики – **квантилі рівня** q **(або** q**-квантилі)**, тобто такі можливі x_q значення ВВ X, для яких $P(X < x_q) = q$.

Деякі квантилі отримали особливі назви. Так, напр., $x_{0,25}$ та $x_{0,75}$ називають відповідно нижнім та верхнім квартилями. Використовують децилі та перцентилі (десяті та соті), а також процентні точки ($100q\%$ точка – це квантиль x_{1-q}, тобто можливе значення ВВ X, при якому $P(X \geq x_{1-q}) = q$).

Розділ 4. Незалежні повторні випробування (НПВ). Формула Бернуллі. Деякі закони розподілу дискретних випадкових величин (Бернуллі, Пуассона, геометричний та гіпергеометричний)

У багатьох практичних задачах доводиться мати справу із серіями випробувань, які проводяться за так званою схемою Бернуллі або схемою незалежних повторних випробувань (НПВ).

Означення. Якщо серію n випробувань проводити в однакових умовах і ймовірність появи події A в кожному окремому випробуванні однакова та не залежить від появи або непояви події A в інших випробуваннях, то таку послідовність НПВ називають схемою Бернуллі.

Прикладами НПВ є: кидки монети (подія A - випадіння цифри), діставання кулі за схемою «повернених куль» із урни з різнокольоровими кулями (подія A - діставання кулі певного кольору), контроль якості серії виготовлених автоматом деталей (подія A - бракована деталь) тощо.

Теорема. Нехай проводиться n НПВ за схемою Бернуллі і ймовірність появи події A (успіху) в кожному із випробувань $p = P(A)$ незмінна (ймовірність непояви події A в кожному із випробувань $q = P(\overline{A}) = 1 - p$). Тоді ймовірність того, що подія A з'явиться m разів у n НПВ $P_n(m) = P_{m,n}$ знаходиться за формулою Бернуллі:

$$P_{m,n} = C_n^m \cdot p^m \cdot q^{n-m}.$$

Доведення.

Зауваження. *При великій кількості НПВ n імовірності появи події m разів зручно обчислювати за допомогою функції «БИНОМРАСП» Excel.*

Приклад. Знайти імовірність того, що при 5 кидках монети герб випаде 3 рази.

Розв'язування. Опишемо стандартну схему НПВ:

$n = 5$ - кількість НПВ (кидки монети);

A - поява герба (подія, що розглядається у кожному окремому випробуванні);

$p = P(A) = 1/2$ - імовірність появи події A;

$q = 1 - p = 1/2$ - імовірність непояви події A;

$m = 3$ - частота (скільки разів) появи події A у n НПВ.

Знайти $P_5(3) = P_{3,5}$.

Розв'язування. Шукану ймовірність знаходимо застосуванням функції *«БИНОМРАСП» Excel* , задаючи її аргументи кількість випробувань п=5, «число успехов» м=3 і «вероятность успеха» р=0,5: = **0,3125**.

Означення. Біноміальным законом розподілу ДВВ X називають ДВВ $X = m$ - частоту появи події A у n НПВ, таблиця розподілу якої має наступний вигляд:

$X = m$	0	1	…	n
$P = P_{m,n}$	$P_{0,n}$	$P_{1,n}$	…	$P_{n,n}$

Відзначимо, що $\sum_{m=0}^{n} P_{m,n} = 1$. Це випливає із формули бінома Ньютона та очевидної рівності $q + p = 1$.

Знайдемо числові характеристики біноміально розподіленої ДВВ $X = m$. Розглянемо випадкові величини $X_i = m_i$, $i = 1,2,..., n$ - частоту появи події A у i -тому випробуванні. Закони розподілу усіх цих ВВ однакові і мають вигляд:

$X_i = m_i$	0	1
P	q	p

Неважко переконатись, що числові характеристики цих ВВ: $M(m_i) = p$, а $D(m_i) = p - p^2 = p(1-p) = pq$. Враховуючи, що $m = \sum_{i=1}^{n} m_i$, за

властивостями математичного сподівання та дисперсії дістанемо:

$$M(m) = M(\sum_{i=1}^{n} m_i) = \sum_{i=1}^{n} M(m_i) = \sum_{i=1}^{n} p = np$$, а

$$D(m) = D(\sum_{i=1}^{n} m_i) = \sum_{i=1}^{n} D(m_i) = \sum_{i=1}^{n} pq = npq$$, звідки

$$\sigma(m) = \sqrt{D(m)} = \sqrt{npq}.$$

Приклад. Статистика стверджує, що 20% пакетів акцій на аукціонах продаються за початково заявленими цінами. Скласти закон розподілу ВВ $X = m$ - частоти проданих за початково заявленими цінами пакетів акцій серед 17 заявлених до торгів. Знайти її числові характеристики.

Розв'язування. Опишемо стандартну схему НПВ:

Приклад показує, що серед значень частоти є таке (в нашому випадку $m_0 = 1$), якому відповідає найбільше значення імовірності.

<u>Означення</u>. **Найімовірнішою частотою** m_0 (або **модою**) появи події A у n НПВ називають частоту, для якої $\forall m: P_{m,n} \le P_{m_o,n}$.

За означенням із системи умов

$$\begin{cases} P_{m_o,n} \ge P_{m_o-1,n} \\ P_{m_o,n} \ge P_{m_o+1,n} \end{cases}$$

неважко дістати подвійну нерівність для визначення найімовірнішої частоти:

$$np - q \le m_o \le np + p.$$

Довжина проміжка, якому належить найімовірніша частота m_0 дорівнює $(np+p)-(np-q) = p+q = 1$, тому m_0 (як ціле число) може приймати або одне значення (якщо кінці проміжка дробові числа), або два значення (якщо кінці проміжка - цілі числа).

У процесі доведення нерівності використовувалось співвідношення $\frac{P_{m+1,n}}{P_{m,n}} = \frac{n-m}{m+1} \cdot \frac{p}{q}$, з якого випливає так звана рекурентна формула Бернуллі:

$$P_{m+1,n} = \frac{n-m}{m+1} \cdot \frac{p}{q} \cdot P_{m,n} .$$

Пропонуємо самостійно переконатись, що ВВ $X = m/n$ - частка (відносна частота) появи події A у n НПВ також підкоряється біноміальному закону розподілу з числовими характеристиками $M(m/n) = p$, $D(m/n) = \frac{pq}{n}$, $\sigma(m/n) = \sqrt{\frac{pq}{n}}$.

Відзначимо, що при достатньо великій кількості випробувань n найімовірніша частота m_0 приблизно дорівнює np, а найімовірніша частка приблизно дорівнює імовірності p появи події (успіху) у кожному окремому випробуванні. Зауважимо також, що біноміальний розподіл при збільшенні кількості НПВ досить швидко наближається до нормального.

Обчислення імовірностей $P_{m,n}$ за точною формулою Бернуллі при великій кількості НПВ n стає досить громіздким , тому на практиці часто використовують так звані асимптотичні формули.

Теорема (локальна формула Муавра-Лапласа). Якщо у схемі Бернуллі із n НПВ імовірність появи події A дорівнює p ($0 < p < 1$), а кількість НПВ досить велика, то імовірність появи події A m разів у n НПВ наближено дорівнює (тим точніше, чим більше n):

$$P_{m,n} \approx \frac{f(x)}{\sqrt{npq}} ,$$

де $f(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$ - функція Гаусса, а $x = \frac{m-np}{\sqrt{npq}}$.

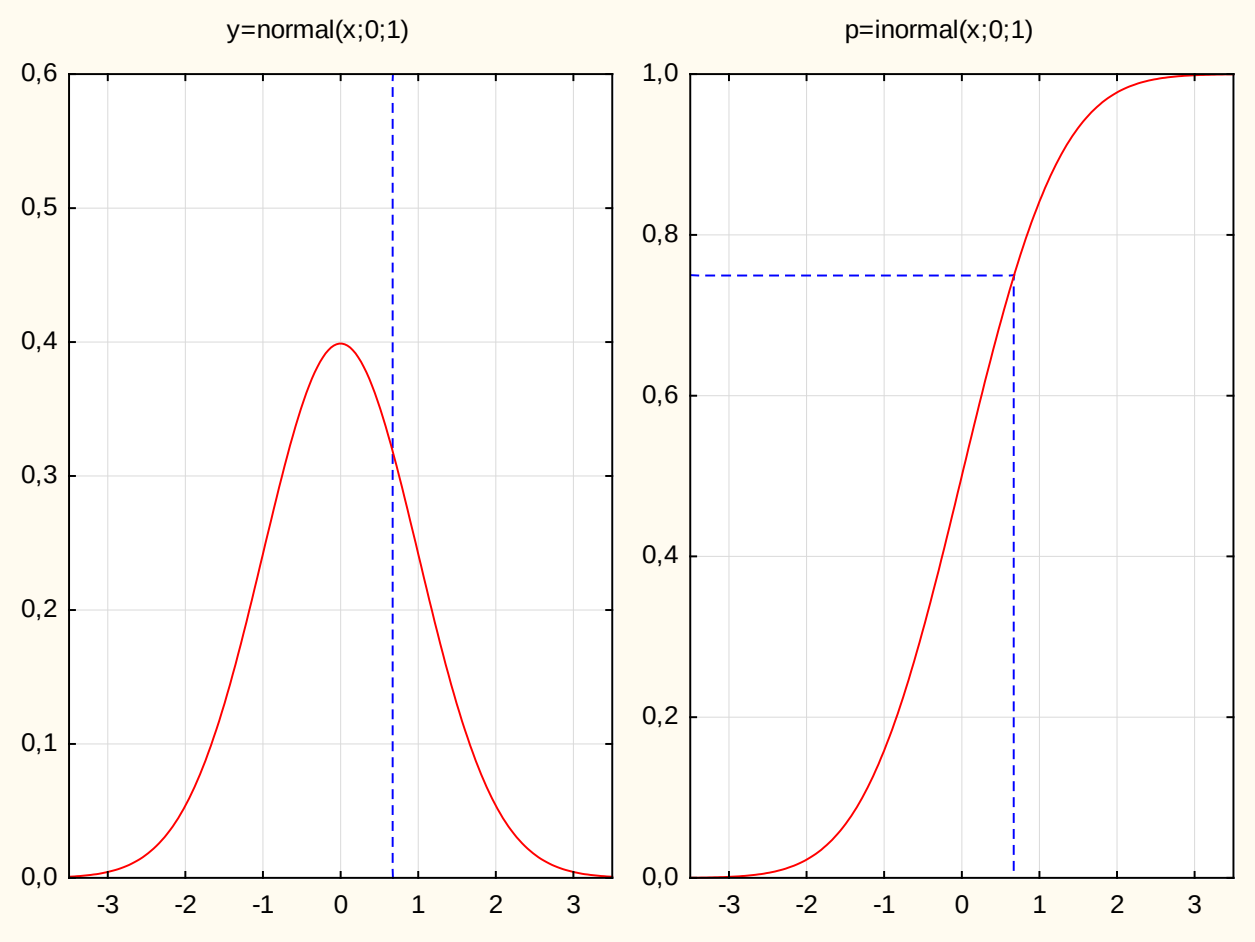
Відзначимо деякі властивості функції Гаусса (локальної функції Лапласа):

1. Функція визначена на всій дійсній осі.
2. Функція набуває тільки додатних значень.

3. Функція парна, тобто $f(-x) = f(x)$.

4. $\lim\limits_{x \to \pm\infty} f(x) = 0$. На практиці при $|x| > 4 \quad f(x) \approx 0$.

Значення цієї функції табульовані для невід'ємних значень x(або користустуються відповідною функцією Excel). Графік функції називають кривою Гаусса або нормальною кривою :



Зауваження. Локальна формула Лапласа дає наближені результати тим ближчі до точних, чим більше значення $\sqrt{npq}$ (при $n \geq 100$, $npq \geq 20$). Це наближення відбувається досить швидко (на практиці формулу застосовують вже навіть при $n > 10$).

Приклад. Монету кидають 25 разів. Знайти найімовірніше число появи герба та його ймовірність.

Розв'язування. Опишемо схему НПВ:

$n = 25$ - кількість НПВ (кидки монети);

А – поява герба при 1 окремому кидку;

$p = P(A) = 0,5;$

$q = 1 - p = 0,5;$

ВВ - частота появи події А (герба) серед п=25 НПВ.

Знайти $m_0 = ?$ - найімовірнішу частоту (моду) появи герба.

Скористуємось подвійною нерівністю: $np - q \le m_0 \le np + p$. Для нашого випадка:

$25 \cdot 0,5 - 0,5 \le m_0 \le 25 \cdot 0,5 + 0,5$. Звідси $m_0 = 12$ або $m_0 = 13$. Відповідні ймовірності знаходимо за локальною формулою Лапласа, використовуючи функцію «НОРМРАСП» із аргументами $a \approx np = 12,5;\ \ \sigma \approx \sqrt{npq} = \sqrt{25 \cdot 0,5 \cdot 0,5} = 2,5$, $x = 12$ або $x = 13$. Відповідні ймовірності дорівнюють: 0,156417 або 0,156417.

Локальна формула Лапласа при малих значеннях $p \le 0,1$ дає досить великі похибки, тому в цих випадках застосовують формулу Пуассона.

Теорема Пуассона. Якщо імовірність появи події A в кожному із випробувань $p \to 0$ при необмеженому зростанні кількості n НПВ ($n \to \infty$), причому добуток np прямує до постійного числа ($np \to a$), то імовірність того, що подія A з'явиться m разів у n НПВ $P_n(m) = P_{m,n}$ задовольняє граничну рівність: $\lim\limits_{n \to \infty} P_{m,n} = \dfrac{a^m e^{-a}}{m!}$.

На практиці, якщо імовірність p постійна і мала, кількість випробувань n - досить велика і число $a = np$ - невелике (при $n \ge 100$, $a = np \le 10$), то користуються наближеною формулою Пуассона: $P_{m,n} \approx \dfrac{a^m e^{-a}}{m!}$.

Формулу називають асимптотичною формулою Пуассона.

Означення. При виконанні умов теореми Пуассона ВВ $X = m$ (яка приймає нескінченну злічену множину значень $m = 0,1,2,...$, а відповідні імовірності знаходяться за формулою $P_m = \dfrac{a^m e^{-a}}{m!}$, де $a = const > 0$) називають **розподіленою за законом Пуассона (закон рідкісних подій).**

Зауваження. *Імовірності* $P_m = \dfrac{a^m e^{-a}}{m!}$ *зручно знаходити за допомогою функції «ПУАССОН» Excel.*

Неважко показати, що для пуассонівського розподілу $M(X) = D(X) = a$, $\sigma(X) = \sqrt{a}$.

Зауважимо, що при малих значеннях p та достатньо великій кількості випробувань n біноміальний розподіл апроксимує пуассонівський.

Приклад. Імовірність виготовлення стандартної деталі дорівнює 0,99. Яка імовірність того, що серед 100 деталей виявиться одна нестандартна?

Розв'язування.

Означення. Течією подій називають послідовність таких подій, які з'являються одна за одною у випадкові моменти часу.

Означення. Течія подій називається **простою або пуассонівською**, якщо вона:

1. **Стаціонарна**, тобто залежить від k - кількості появ події за проміжок часу t і не залежить від моменту початку цього проміжку;
2. Має властивість **відсутності післядії**, тобто імовірність появи події не залежить від її появи або непояви раніше та не впливає на майбутнє;
3. **Ординарна**, тобто імовірність появи більше однієї події за малий проміжок часу є величина нескінченно мала.

Означення. Середнє число λ появ події за одиницю часу називають **інтенсивністю течії**.

Теорема. Якщо течія пуассонівська (проста), то імовірність появи події k разів за час t можна знайти за формулою (математичною моделлю простої течії подій):

$$P_t(k) = \frac{(\lambda t)^k e^{-\lambda t}}{k!},$$

де λ - інтенсивність течії.

Прикладами простої течії подій можуть бути: поява викликів на АТС, прибуття літаків до аеропорту, прихід покупців до супермаркету тощо.

Означення. Нехай ДВВ $X = m$ - кількість випробувань до появи події A в серії НПВ, яка може приймати нескінченну злічену множину значень $m = 1,2,3,...$, а відповідні ймовірності знаходяться за формулою

$$P(X = m) = pq^{m-1},$$

де $p = P(A)$ - імовірність появи події A в кожному випробуванні, $q = 1 - p$. Така ДВВ називається розподіленою за **геометричним законом**.

Ряд відповідних імовірностей цього розподілу є нескінченно спадною геометричною прогресією зі знаменником q, сума якого дорівнює одиниці (умова нормування).

Неважко показати, що для геометрично розподіленої ВВ:

$$M(X) = \frac{1}{p} \quad , \quad D(X) = \frac{q}{p^2} \quad , \quad \sigma(X) = \frac{\sqrt{q}}{p}.$$

Геометричний розподіл застосовується у різноманітних задачах статистичного контролю якості виробів, в теорії надійності та в страхових розрахунках.

Означення. Нехай ДВВ $X = m$ - кількість елементів із певною властивістю серед n елементів, відібраних із сукупності в N елементів, яка містить k, $k \geq n$ елементів саме такої властивості. Ця ДВВ може приймати значення $m = 0,1,2,\ldots,n$ з імовірностями $P(X = m) = \frac{C_k^m \cdot C_{N-k}^{n-m}}{C_N^n}$ і підкоряється **гіпергеометричному закону розподілу**.

Для цієї ДВВ $M(X) = \frac{kn}{N}$, $D(X) = \frac{nk(N-k)(N-n)}{N^2(N-1)}$.

Гіпергеометричний розподіл використовують у багатьох задачах статистичного контролю якості.

Відзначимо, що при малих об'ємах вибірки n у порівнянні із об'ємом N усієї сукупності ($\frac{n}{N} \leq 0,1$; $\frac{n}{k} \leq 0,1$; $\frac{n}{N-k} \leq 0,1$) імовірності у гіпергеометричному розподілі будуть близькими до відповідних імовірностей біноміального розподілу з $p = \frac{k}{N}$. В статистиці це означає, що розрахунки імовірностей для безповторної вибірки будуть мало відрізнятись від таких розрахунків для повторної вибірки.

Розділ 5. Функції розподілу випадкових величин. Неперервні випадкові величини (НВВ). Закони розподілу деяких НВВ (рівномірний, показниковий, нормальний). Деякі розподіли, пов'язані з нормальним

Існує універсальний спосіб задання ВВ – за допомогою функції розподілу імовірностей (або інтегральної функції розподілу). Всюди надалі ВВ позначаються великими літерами, а малими $x, y, z, \ldots$ - довільні дійсні числа.

Означення. Інтегральною функцією розподілу $F(x)$ ВВ X (або просто – функцією розподілу) називається ймовірність того, що ВВ прийме значення, менше від числа x, тобто

$$F(x) = P(X < x) = P(-\infty < X < x).$$

Ця функція повністю характеризує ВВ з імовірнісної точки зору, тобто є однією із форм закону розподілу. Тепер можна дати чітке означення ДВВ та НВВ.

Означення. Випадковою величиною називається величина, яка може приймати значення в залежності від випадкових обставин і для якої визначена функція розподілу ймовірностей. ВВ називається **неперервною (НВВ)**, якщо її інтегральна функція (функція розподілу) - неперервна. ВВ називається **дискретною (ДВВ)**, якщо її функція розподілу - розривна (кусочно – стала).

Для ДВВ X із множиною значень $x_1, \ldots, x_n, \ldots$ функція розподілу ймовірностей визначається як

$$F(x) = P(X < x) = \sum_{x_i < x} P(X = x_i),$$

де символ $x_i < x$ означає, що сумування проводиться для всіх можливих значень x_i, які менші від x.

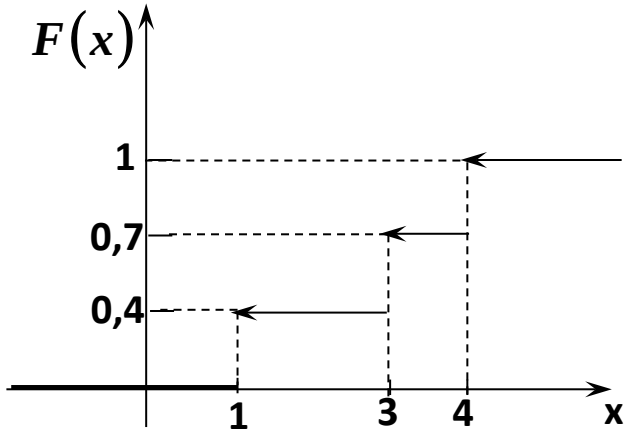
Приклад. Таблиця розподілу ДВВ X має наступний вигляд:

X	1	3	4
P	0,4	0,3	0,3

Знайти інтегральну функцію розподілу, побудувати її графік.

Розв'язування дає наступний результат

Графік інтегральної функції розподілу ВВ X зображений на рис.



Її аналітичний вираз:

$$F(x)=\begin{cases} 0, & x\le 1,\\ 0{,}4, & 1<x\le 3,\\ 0{,}7, & 3<x\le 4,\\ 1, & x>4. \end{cases}$$

Висновок: функція розподілу ДВВ розривна (кусочно-стала), причому розриви типу «стрибків» відбуваються у тих точках, у яких ДВВ набуває можливих значень. Зазначимо, що по вигляду цієї функції (або по її графіку) легко відновити таблицю розподілу.

ВЛАСТИВОСТІ ФУНКЦІЇ РОЗПОДІЛУ.

Всюди надалі вважається, що інтегральна функція визначена $\forall x\in R$.

1. Значення функції належать проміжку $[0;1]$, тобто $0\le F(x)\le 1$, причому $F(-\infty)=\lim\limits_{x\to-\infty}F(x)=0$, $F(+\infty)=\lim\limits_{x\to+\infty}F(x)=1$.

Доведення.

2. Функція є неспадною, тобто $\forall x_1<x_2:\quad F(x_1)\le F(x_2)$.

Доведення.

Наслідок (основна формула теорії ймовірностей) :

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

3. Імовірність того, що НВВ X прийме деяке окреме значення дорівнює нулю, тобто $P(X=x_1)=0$.

Доведення.

Наслідок . Для НВВ X справедливі рівності: $P(x_1 \le X < x_2) = P(x_1 \le X \le x_2) = P(x_1 < X < x_2) = P(x_1 < X \le x_2)$ тощо

Розглянуті властивості функцій розподілу можна сформулювати наступним чином: будь-яка функція розподілу є невід'ємною неспадною функцією, що задовольняє умови $F(-\infty) = 0, \quad F(+\infty) = 1$. Справедливе і обернене твердження: будь-яка функція, що задовольняє вищевказаним властивостям, може бути функцією розподілу деякої ВВ.

Приклад. Нехай річний дохід навмання вибраного підприємця є ВВ X, розподіленою за законом Парето з параметрами $\alpha > 0$ та $x_o > 0$ (x_o - граничний дохід, що не обкладається податком). Функція розподілу ВВ X має наступний вигляд:

$$F(x) = \begin{cases} 0 \ , & x \le x_o \\ 1 - (x_o / x)^{\alpha} \ , & x > x_o \end{cases}$$

Побудувати графік функції розподілу при $\alpha = 2$ та визначити розмір річного доходу підприємця, що обкладається податком, який може бути перевищений з імовірністю 0,5 .

Розв'язування.

ЩІЛЬНІСТЬ РОЗПОДІЛУ ЙМОВІРНОСТЕЙ.

Незважаючи на те, що інтегральна функція розподілу повністю характеризує ВВ, вона не дозволяє уявити характер розподілу. Тому користуються (тільки для НВВ) так званою диференціальною функцією або щільністю розподілу ймовірностей, яка також є законом розподілу.

Означення. **Щільністю розподілу ймовірностей (або диференціальною функцією розподілу)** називається похідна (якщо вона існує) від інтегральної функції розподілу:

$$\varphi(x) = F'(x) = \lim_{\Delta x \to 0} \frac{\Delta F(x)}{\Delta x} = \lim_{\Delta x \to 0} \frac{F(x + \Delta x) - F(x)}{\Delta x}.$$

ВЛАСТИВОСТІ ДИФЕРЕНЦІАЛЬНОЇ ФУНКЦІЇ.

1. Щільність розподілу імовірностей – невід'ємна функція, тобто $\varphi(x) \geq 0$.
2. **Теорема** (основна формула теорії імовірностей). Імовірність того, що НВВ X прийме значення із деякого проміжка $(a;b)$ знаходиться за формулою:

$$P(a < X < b) = \int_a^b \varphi(x)dx.$$

3. Інтегральна функція розподілу та диференціальна (щільність розподілу імовірностей) є еквівалентними узагальненими характеристиками НВВ, які пов'язані співвідношенням: $F(x) = \int_{-\infty}^{x} \varphi(t)dt.$

4. Умова нормування закону розподілу НВВ має вигляд: $\int_{-\infty}^{+\infty} \varphi(x)dx = 1.$

Числові характеристики НВВ X визначаються наступними формулами (якщо збігаються відповідні невласні інтеграли):

$$M(X) = \int_{-\infty}^{+\infty} x\varphi(x)dx\,,$$

$$D(X) = \int_{-\infty}^{+\infty} [x - M(X)]^2 \varphi(x)dx = M(X^2) - M^2(X)\,,$$

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

ЗАКОНИ РОЗПОДІЛУ ДЕЯКИХ НВВ.

РІВНОМІРНИЙ РОЗПОДІЛ

Означення. НВВ X називається **рівномірно розподіленою на проміжку** $[a;b]$, якщо її щільність розподілу імовірностей стала на цьому проміжку, а поза цим проміжком дорівнює нулю, тобто

$$\varphi(x) = \begin{cases} c, & x \in [a;b] \\ 0, & x \notin [a;b] \end{cases}$$

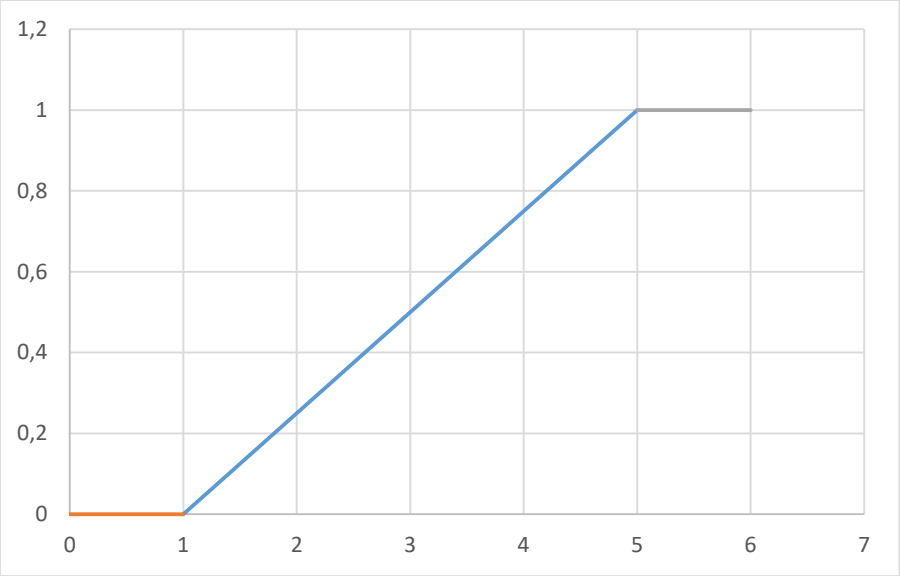
Знайдемо значення сталої c, скориставшись властивістю диференціальної функції (умовою нормування):

$$\int_{-\infty}^{+\infty}\varphi(x)dx=1,\quad \int_{-\infty}^{a}0\cdot dx+\int_{a}^{b}c\cdot dx+\int_{b}^{+\infty}0\cdot dx=1,\quad c(b-a)=1.$$

Звідси $c=\frac{1}{b-a}$. Таким чином, щільність рівномірно розподіленої ВВ має вигляд:

$$\varphi(x)=\begin{cases}\frac{1}{b-a}, & x\in[a;b]\\ 0, & x\notin[a;b]\end{cases}$$

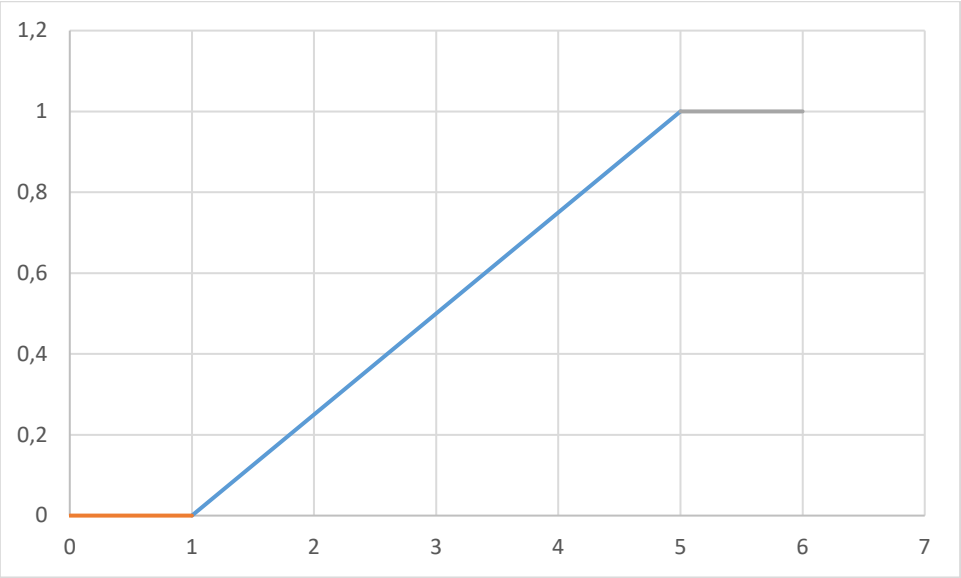
Її графік (наприклад, при a=1; b=5):



Функція розподілу рівномірно розподіленої НВВ X:

$$F(x)=\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & a\le x\le b\\ 1, & x>b\end{cases}$$

Графік функції розподілу (при a=1; b=5) зображений на рис.



Числові характеристики рівномірно розподіленої НВВ X визначаються формулами: $M(X)=\frac{a+b}{2}$, $D(X)=\frac{(b-a)^2}{12}$, $\sigma(X)=\frac{b-a}{2\sqrt{3}}$.

Зазначимо, що розподіл суми незалежних рівномірно розподілених на $[0;1]$ НВВ : $X=X_1+X_2+\ldots+X_n$, нормованих математичним сподіванням $n/2$ та середнім квадратичним відхиленням $\sqrt{n/12}$, зі зростанням n швидко прямує до стандартного нормального розподілу (вже при $n\geq 3$).

ПОКАЗНИКОВИЙ (ЕКСПОНЕНЦІЙНИЙ) РОЗПОДІЛ

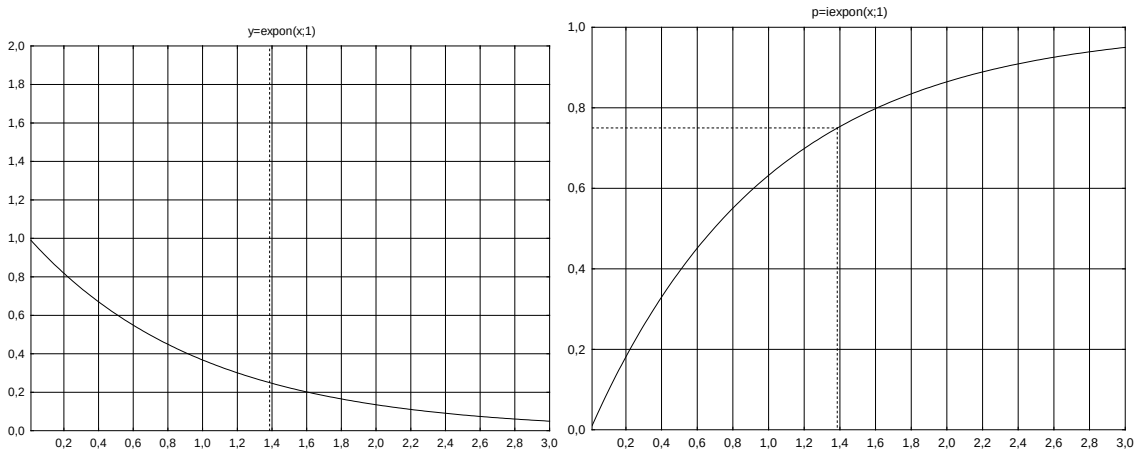
Означення. НВВ X називається **розподіленою за показниковим законом з параметром** $\lambda>0$, якщо її щільність розподілу імовірностей має вигляд

$$\varphi(x)=\begin{cases}0, & x<0\\ \lambda e^{-\lambda x}, & x\geq 0\end{cases}$$

Неважко впевнитись, що інтегральна функція розподілу для НВВ X, розподіленої за показниковим розподілом, має вигляд:

$$F(x)=\begin{cases}0, & x<0\\ 1-e^{-\lambda x}, & x\geq 0\end{cases}$$

Графіки цих функцій (при «лямбда»=2):



Числові характеристики для НВВ X , розподіленої за показниковим розподілом, визначаються формулами: $M(X)=\frac{1}{\lambda}$, $D(X)=\frac{1}{\lambda^2}$, $\sigma(X)=\frac{1}{\lambda}$.

Зауважимо, що показниковий розподіл широко застосовується в теорії надійності та в системах масового обслуговування тощо. Зокрема, тільки цьому закону підкоряється час між появою двох послідовних подій у найпростішій течії подій.

НОРМАЛЬНИЙ ЗАКОН РОЗПОДІЛУ

Означення. НВВ X розподілена за нормальним законом з параметрами a та $\sigma>0$, якщо її щільність розподілу імовірностей має вигляд:

$$\varphi(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}} .$$

Скористувавшись означеннями, неважко переконатись, що числові характеристики нормально розподіленої ВВ дорівнюють:

$$M(X)=a \quad , \quad D(X)=\sigma^2 \quad , \quad \sigma(X)=\sigma .$$

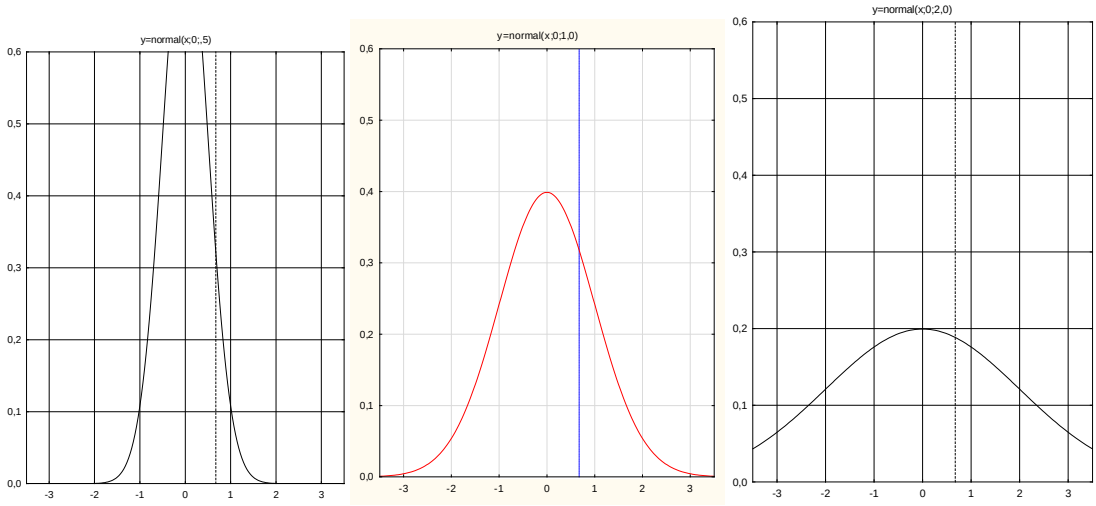
Властивості диференціальної функції – щільності $\varphi(x)$ нормально розподіленої ВВ:

1. Функція визначена і неперервна на всій числовій осі.
2. Функція набуває лише додатних значень.
3. Графік симетричний відносно прямої $x=a$.

4. Точкою максимума є $x_o = a$, причому $\varphi_{\max}(x) = \varphi(a) = \dfrac{1}{\sigma\sqrt{2\pi}}$.

5. $\lim\limits_{x \to \pm\infty} \varphi(x) = 0$, тобто вісь абсцис є асимптотою графіка функції.

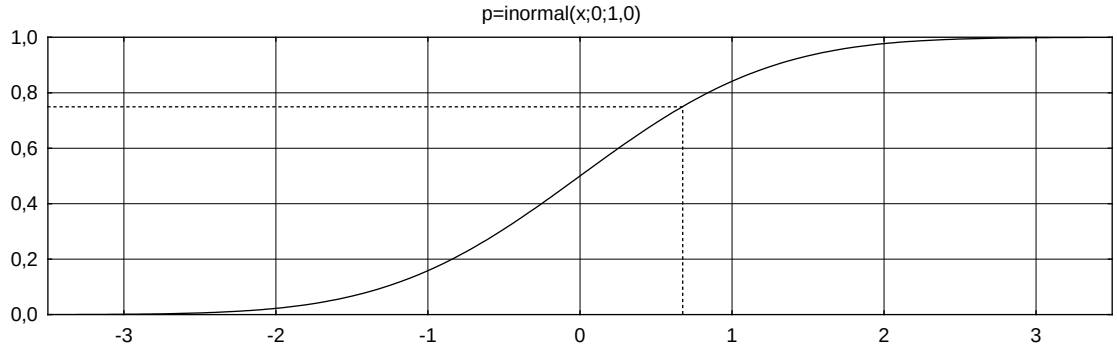
Характер поведінки функції в залежності від параметрів розподілу показані на рисунках (a=0; “sigma”=0,5; 1; 2):



За властивостями (співвідношення між функціями розподілу) можна знайти вигляд функції розподілу (інтегральної):

$$F(x) = \int_{-\infty}^{x} \varphi(t)dt = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-a)^2}{2\sigma^2}} dt.$$

Графік інтегральної функції (a=0; «сигма»=1):



Знайдемо вигляд інтегральної функції розподілу. За властивостями (співвідношення між функціями розподілу):

$$F(x)=\int_{-\infty}^{x}\varphi(t)dt=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(t-a)^2}{2\sigma^2}}dt.$$

Зробимо заміну змінних, поклавши $\frac{x-a}{\sigma}=z \quad , \quad x=a+\sigma z \quad , \quad dx=\sigma dt$:

Враховуючи інтеграл Пуассона $\int_{-\infty}^{+\infty}e^{-\frac{z^2}{2}}dz=\sqrt{2\pi}$ та використавши інтегральну функцію Лапласа $\Phi(x)=\frac{1}{\sqrt{2\pi}}\int_{0}^{x}e^{-\frac{t^2}{2}}dt$, дістаємо:

$$F(x)=\frac{1}{2}+\Phi(\frac{x-a}{\sigma}) .$$

При дослідженні інтегральної функції $F(x)$ враховані наступні властивості інтегральної функції Лапласа $\Phi(x)$:

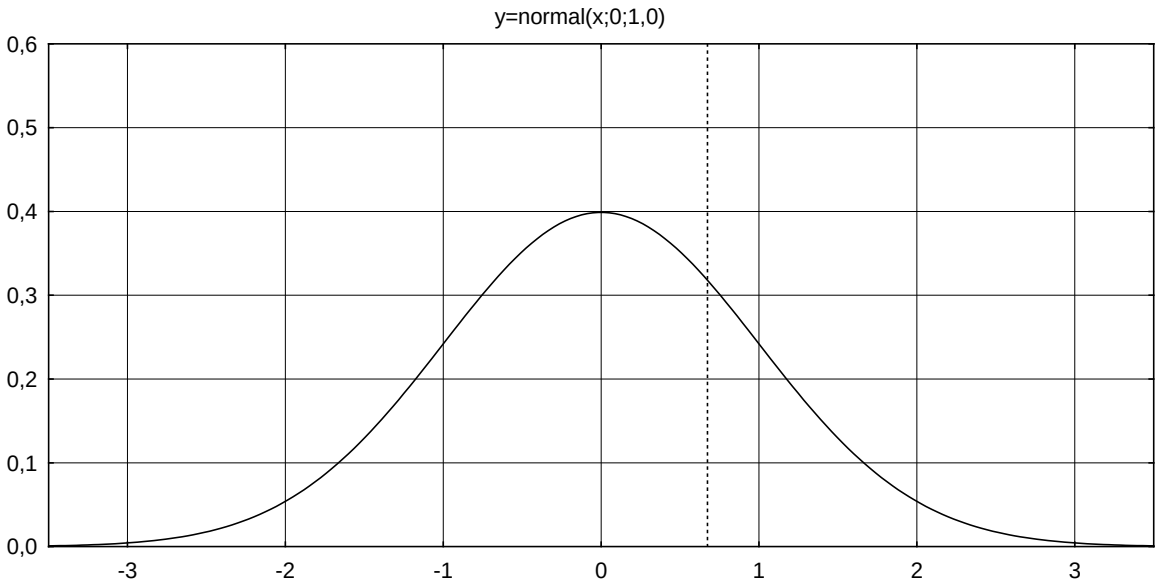
1) функція визначена та неперервна на всій числовій осі;
2) $\Phi(0)=0$;
3) функція непарна ($\Phi(-x)=\Phi(x)$) , тому її графік симетричний відносно початку координат;
4) функція монотонно зростає, причому $\lim\limits_{x\to-\infty}\Phi(x)=-1/2 \quad , \quad \lim\limits_{x\to+\infty}\Phi(x)=1/2.$

Зазначимо, що значення функції Лапласа табульовані при $x\in[0;5]$ *, а при* $x>5 \quad \Phi(x)\approx 1/2$. *Але набагато ефективніше використовувати функцію «НОРМРАСП» Excel, яка дозволяє обчислювати значення диференціальної та інтегральної функцій нормального розподілу.*

ВЛАСТИВОСТІ НОРМАЛЬНОГО РОЗПОДІЛУ.

Правило трьох сигм. Із практичною достовірністю (з імовірністю 0,9973) можна стверджувати, що значення нормально розподіленої ВВ X попадають до проміжка $[a-3\sigma; a+3\sigma]$.

Графічно це можна зобразити так (при а=0; «сигма»=1):



Розглянуті властивості дозволяють виділити три характерних особливості нормально розподіленої ВВ X:
а) найчастіше у розподілі зустрічаються значення ВВ, близькі до середнього;
б) значення ВВ, рівновіддалені від середнього, зустрічаються у розподілі однаково часто;
в) по мірі віддалення значень ВВ від центру розподілу вони зустрічаються все рідше та рідше.

Досить часто на практиці довільну ВВ X, розподілену за нормальним законом, нормують, тобто, замість неї розглядають **стандартизовану нормально розподілену ВВ** $Z=\dfrac{X-a}{\sigma}$ **, числові характеристики якої дорівнюють** $M(Z)=0 \quad , \quad \sigma(Z)=1.$

ДЕЯКІ РОЗПОДІЛИ, ПОВ'ЯЗАНІ З НОРМАЛЬНИМ.

Логнормальний розподіл.

Неперервна ВВ X, яка набуває додатних значень, підкоряється **логарифмічно нормальному закону розподілу** (скорочено – **логнормально розподілена**), якщо її логарифм $\ln X$ є нормально розподіленою ВВ.

Розподіл визначається двома параметрами a та σ, причому, якщо у нормального розподіла a - це середнє значення (математичне сподівання), то для логнормального розподіла параметр a - це його медіана. Логнормальний розподіл суттєво асиметричний (крива щільності круто підіймається зліва від a і полого спускається справа). При прямуванні σ до нуля логнормальний розподіл прямує до нормального.

Логнормальний розподіл виникає у моделях росту, часто використовується для описування розподілу доходів, банківських вкладів, місячної зарплати, дебіту нафтових свердловин, посівних площ, довговічності виробів тощо.

χ^2 - розподіл.

Розподілом χ^2 («хі-квадрат») із k ступенями вільності називається розподіл суми квадратів k незалежних ВВ, розподілених за стандартним нормальним законом, тобто

$$\chi^2 = \sum_{i=1}^{k} Z_i ,$$

де $Z_i, i = 1,2,...,k$ підкоряються стандартному нормальному закону (з нульовими математичними сподіваннями і середньоквадратичними відхиленнями, рівними одиниці). χ^2 - розподіл має правосторонню асиметрію і при зростанні k повільно прямує до нормального. Цей розподіл широко застосовується в математичній статистиці.

Розподіл Стьюдента.

Розподілом Стьюдента (t-розподілом) з k ступенями вільності називається розподіл ВВ

$$t = \frac{Z}{\sqrt{\frac{\chi^2}{k}}},$$

де Z - ВВ, розподілена за стандартним нормальним законом, а χ^2 - незалежна від Z ВВ, яка має χ^2-розподіл із k степенями вільності. Розподіл Стьюдента близький до нормального, але більш пологий (із більш довгими «хвостами»). При зростанні k він швидко наближається до нормального розподілу. Розподіл Стьюдента широко застосовується в математичній статистиці.

Група теорем, які встановлюють відповідність між теоретичними та експериментальними характеристиками великої кількості випадкових величин і випадкових подій, а також які стосуються граничних законів розподілу, об'єднуються під загальною назвою граничних теорем теорії ймовірностей. Ці теореми поділимо на дві групи: закон великих чисел та центральну граничну терему. За А.Н.Колмогоровим під законом великих чисел розумітимемо загальний принцип, згідно до якого сукупна дія великої кількості випадкових факторів призводить (при деяких досить загальних умовах) до результату, який майже не залежить від випадку. Іншими словами, при великій кількості ВВ їх середній результат втрачає випадковість і може бути передбаченим із великою ступінню визначеності.

Теорема (нерівність Маркова). Якщо ВВ X приймає тільки невід'ємні значення і має фіксоване математичне сподівання $M(X)$, то для довільного додатного числа α справедлива нерівність:

$$P(X \geq \alpha) \leq \frac{M(X)}{\alpha}.$$

Наслідок (друга форма нерівності Маркова) . За умовами теореми

$$P(X < \alpha) > 1 - \frac{M(X)}{\alpha}.$$

Приклад. Банк обслуговує в середньому 100 клієнтів щодня. Оцінити імовірність того, що деякого дня банк обслугує: а) не менше 200 клієнтів; б) менше 150 клієнтів.

Розв'язування.

Приклад. Сума всіх вкладів населення у деякому банку становить 3млн.грн., а імовірність того, що випадково взятий вклад буде меншим від 10тис.грн., дорівнює 0,8 . Що можна сказати про кількість вкладчиків банку?

Розв'язування.

Теорема (нерівність Чебишова). Якщо довільна ВВ X має фіксовані математичне сподівання $M(X)$ та дисперсію $D(X)$, то для довільного додатного числа ε справедлива нерівність:

$$P\left(|X - M(X)| \geq \varepsilon\right) \leq \frac{D(X)}{\varepsilon^2}.$$

Наслідок (друга форма нерівності Чебишова) . За умовами теореми

$$P\left(\left|X - M(X)\right| < \varepsilon\right) > 1 - \frac{D(X)}{\varepsilon^2} .$$

Зауваження. Нерівності Маркова, Чебишова дають оцінки імовірностей подій знизу або зверху. Часто ці оцінки занадто грубі, а інколи просто тривіальні.

ЧАСТИННІ ВИПАДКИ НЕРІВНОСТІ ЧЕБИШОВА.

а) для біноміально розподіленої ДВВ $X = m$ - частоти появи події A з імовірністю p в серії із n НПВ:

$$P\left(\left|m - np\right| < \varepsilon\right) > 1 - \frac{npq}{\varepsilon^2} \text{ , або } P\left(\left|m - np\right| \geq \varepsilon\right) \leq \frac{npq}{\varepsilon^2} ;$$

б) для біноміально розподіленої ДВВ $X = \frac{m}{n}$ - частості (частки) появи події A в серії із n НПВ:

$$P\left(\left|\frac{m}{n} - p\right| < \varepsilon\right) > 1 - \frac{pq}{n\varepsilon^2} \text{ , або } P\left(\left|\frac{m}{n} - p\right| \geq \varepsilon\right) \leq \frac{pq}{n\varepsilon^2} .$$

Приклад. Середньодобове споживання води мешканцем Одеси становить 300л, а середнє квадратичне відхилення не перевищує 60л. Оцінити імовірність того, що у випадково обрану добу споживання води мешканцем менше 600л.

Розв'язування.

Теорема Чебишова. Якщо всі дисперсії послідовності попарно незалежних ВВ $X_1, \ldots, X_n, \ldots$ не перевищують деякого додатного числа, то при $n \to \infty$ майже достовірним можна вважати подію, яка полягає у тому, що модуль відхилення середнього арифметичного ВВ від середнього арифметичного їх математичних сподівань буде величиною нескінченно малою, тобто:

$$\lim_{n\to\infty} P\left(\left|\frac{X_1+\ldots+X_n}{n}-\frac{a_1+\ldots+a_n}{n}\right|\le\varepsilon\right)=1.$$

Сенс теореми Чебишова полягає у тому, що при великій кількості незалежних ВВ, дисперсії яких обмежені у сукупності, їх середня арифметична практично втрачає характер ВВ і як завгодно мало відрізняється від сталої величини.

Означення. Послідовність ВВ $X_1,\ldots,X_n,\ldots$ називається збіжною за імовірністю до величини a (сталої або випадкової), якщо для довільного як завгодно малого числа ε

$$\lim_{n\to\infty} P\left(|X_n-a|\le\varepsilon\right)=1.$$

Часто застосовується позначення $X_n \underset{n\to\infty}{\overset{імов.}{\to}} a$.

Теорема Бернуллі. Частка $\frac{m}{n}$ появи події A в серії із n НПВ при $n\to\infty$ збігається за імовірністю до $p=P(A)$ - імовірності появи події у кожному окремому випробуванні:

$$\lim_{n\to\infty} P\left(\left|\frac{m}{n}-p\right|<\varepsilon\right)=1 \quad \forall\varepsilon>0 \text{ , або } \frac{m}{n} \underset{n\to\infty}{\overset{імов.}{\to}} p .$$

Теорема Ляпунова. Якщо $X_1,\ldots,X_n,\ldots$ - незалежні ВВ, у кожної із яких існують математичні сподівання $a_i=M(X_i)$, дисперсії $\sigma_i=D(X_i)$ та абсолютні центральні моменти третього порядка $m_i=M\left(|X_i-a_i|^3\right)$, причому виконується умова $\lim_{n\to\infty}\frac{\sum_{i=1}^{n} m_i}{\sum_{i=1}^{n}\sigma^2{}_i}=0$.

Тоді закон розподілу суми $X_1+\ldots+X_n$ при $n\to\infty$ необмежено

наближається до нормального з математичним сподіванням $a = \sum_{i=1}^{n} a_i$ і дисперсією $\sigma^2 = \sum_{i=1}^{n} \sigma^2{}_i$.

<u>Наслідок (центральна гранична теорема)</u>. Якщо всі ВВ однаково розподілені, то закон розподілу їх суми при $n \to \infty$ необмежено наближається до нормального.

<u>Інтегральна теорема Муавра-Лапласа</u>. Для біноміально розподіленої ДВВ $X = m$ - частоти появи події A з імовірністю p в серії із n НПВ справедлива наближена формула:

$$P(m_1 \le m \le m_2) \approx \Phi(t_2) - \Phi(t_1) ,$$

де $\Phi(t)$ - інтегральна функція нормального розподілу, $t_1 = \frac{m_1 - np}{\sqrt{npq}}$, $t_2 = \frac{m_2 - np}{\sqrt{npq}}$.

<u>Частинні випадки інтегральної теореми Муавра-Лапласа</u> . Для частоти m та частості $\frac{m}{n}$ появи події A з імовірністю p в серії із n НПВ справедливі наближені формули:

$$P\left(|m - np| < \varepsilon\right) \approx 2\Phi\left(\frac{\varepsilon}{\sqrt{npq}}\right) ,$$

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

РОЗДІЛ 7. СИСТЕМА ВИПАДКОВИХ ВЕЛИЧИН

Раніше розглядались ВВ, які при кожному випробуванні визначались одним можливим значенням. Тому таку ВВ називають **одновимірною**.

Якщо можливі значення ВВ визначаються у кожному випробуванні $2,3,\ldots,n$ числами, то такі ВВ називають відповідно **дво-**, **трьох-**,..., n**-вимірними**. Двовимірну ВВ будемо позначати (X,Y), де X та Y - компоненти. ВВ X та Y, що розглядаються одночасно, утворюють систему двох випадкових величин. Аналогічно можна розглядати систему n ВВ.

Означення. Сукупність n одночасно розглянутих ВВ $(X_1,\ldots,X_n)$ називають **системою ВВ**.

Систему n ВВ $(X_1,\ldots,X_n)$ можна розглядати як випадкову точку в n-вимірному просторі з координатами $(x_1,\ldots,x_n)$ або як випадковий вектор, напрямлений з початку координат у точку $A(x_1,\ldots,x_n)$.

При $n=2$ дістаємо систему двох ВВ (X,Y), яку можна інтерпретувати як випадкову точку $A(x,y)$ на площині xOy або як випадковий вектор $\overline{OA}$:

Багатовимірні ВВ можуть бути **дискретними** - ДВВ або **неперервними** –НВВ (компоненти цих величин відповідно будуть дискретними або неперервними).

Означення. **Законом розподілу двовимірної ДВВ** називають сукупність із множини її можливих значень (x_i,y_j) та їх імовірностей $p_{ij}=P(X=x_i,Y=y_j)$, $i=1,2,\ldots,n;\ j=1,2,\ldots,m$.

Найчастіше закон розподілу двовимірної ДВВ задають таблицею:

X	Y					
	y_1	y_2	…	y_j	…	y_m
x_1	p_{11}	p_{12}	…	p_{1j}	…	p_{1m}
x_2	p_{21}	p_{22}	…	p_{2j}	…	p_{2m}
…	…	…	…	…	…	…
x_i	p_{i1}	p_{i2}	…	p_{ij}	…	p_{im}
…	…	…	…	…	…	…
x_n	p_{n1}	p_{n2}	…	p_{nj}	…	p_{nm}

Події $(X = x_i, Y = y_j)$., $i = 1,..., n;\ j = 1,..., m$ утворюють повну групу, тому сума імовірностей таблиці дорівнює одиниці, тобто виконується умова нормування: $\sum_{i=1}^{n} \sum_{j=1}^{m} p_{ij} = 1$.

Закон розподілу двовимірної ДВВ дозволяє отримати закони розподілу кожної компоненти. Відповідні імовірності для можливих значень компонент знаходяться сумуванням імовірностей у рядках та стовпцях таблиці.

Приклад. Знайти закони розподілу компонент двовимірної ВВ, закон розподілу якої заданий таблицею:

X	Y	
	y_1	y_2
x_1	0,1	0,06
x_2	0,3	0,18
x_3	0,2	0,16

Розв'язування:

Означення. Інтегральною функцією розподілу (функцією розподілу) двовимірної ВВ (X,Y) називають функцію двох змінних $F(x,y)$, яка визначає для кожної пари (X,Y) імовірність виконання нерівностей $X<x;Y<y$, тобто

$$F(x;y)=P(X<x;Y<y).$$

Геометричний сенс функції розподілу $F(x,y)$ - це імовірність того, що випадкова точка $A(X,Y)$ попаде у нескінченний прямокутник з вершиною в точці (x,y):

Неважко переконатись у наступних властивостях функції розподілу:

1) $0\le F(x,y)\le 1$, причому $F(-\infty,y)=F(x,-\infty)=F(-\infty,+\infty)=0;\quad F(+\infty,+\infty)=1.$

2) $F(x,y)$ - неспадна функція за кожним аргументом, тобто

$F(x_1,y)\ge F(x_2,y)$, якщо $x_1>x_2$;

$F(x,y_1)\ge F(x,y_2)$, якщо $y_1>y_2$.

3) Імовірність попадання випадкової точки до прямокутника $\{x_1\le X<x_2;y_1\le Y<y_2\}$ можна знайти за формулою:

$$P(x_1\le X<x_2;y_1\le Y<y_2)=$$
$$=\{F(x_2,y_2)-F(x_1,y_2)\}-\{F(x_2,y_1)-F(x_1,y_1)\}$$

Зауважимо, що інтегральна функція розподілу ДВВ – розривна (кусочно-стала), а у неперервних ВВ – неперервна. НВВ можна задавати також щільністю імовірностей.

Означення. Диференціальною функцією розподілу (двовимірною щільністю імовірностей) $f(x,y)$ двовимірної НВВ (X,Y) називають мішану частинну похідну другого порядку від інтегральної функції розподілу

$$f(x,y)=\frac{\partial^2 F(x,y)}{\partial x\partial y}=F''_{xy}.$$

Щільність розподілу імовірностей $f(x,y)$ задовольняє властивостям:

1) Вона невід'ємна, тобто $f(x,y)\ge 0$ (як похідна неспадної функції).

2) $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x,y)dxdy = 1$ (умова нормування).

3) $F(x,y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(t,s)dtds$ (зв'язок із інтегральною функцією).

4) Імовірність попадання випадкової точки $A(X,Y)$ в область D знаходиться за формулою $P\big(A(X,Y) \in D\big) = \iint\limits_{D} f(x,y)dxdy$.

Дві випадкові величини **незалежні**, якщо закон розподілу однієї з них не залежить від того, які можливі значення прийняла інша величина. У супротивному випадку випадкові величини **залежні**.

Теорема. Для того, щоб ВВ X та Y були незалежними, необхідно і достатньо, щоб інтегральна (або диференціальна) функція системи (X,Y) дорівнювала добутку інтегральних (диференціальних) функцій компонент

$$F(x,y) = F_1(x)\cdot F_2(y) \text{ (або } f(x,y) = f_1(x)\cdot f_2(y)\text{).}$$

Наслідок. Для незалежності двох ДВВ необхідно і достатньо, щоб $p_{ij} = P(X = x_i, Y = y_j) = P(X = x_i)\cdot P(Y = y_j)$.

Числові характеристики двохвимірної ВВ.

Математичне сподівання двохвимірної (системи) ВВ (X,Y) позначається $(\overline{X},\overline{Y})$ характеризує координати **центру розподілу** ВВ. Ці координати у випадку НВВ знаходяться за формулами:

$$\overline{X} = m_X = M(X) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} xf(x,y)dxdy,$$

$$\overline{Y} = m_Y = M(Y) = \int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} yf(x,y)dxdy.$$

Дисперсії D_X та D_Y характеризують розсіювання випадкової точки (X,Y) від центру розподілу $(\overline{X},\overline{Y})$ вздовж координатних осей Ох та Оу відповідно. Їх можна знаходити за формулами:

$$D_X = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^2 f(x,y)dxdy - m_X^2,$$

$$D_Y = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y^2 f(x,y)dxdy - m_Y^2.$$

Для опису двовимірної ВВ крім математичного сподівання, дисперсії та середніх квадратичних відхилень $\sigma_X = \sqrt{D_X}$, $\sigma_Y = \sqrt{D_Y}$ використовують також **кореляційний момент (коваріацію):**

$$\mathrm{cov}(X,Y) = K_{XY} = M\left[(X - m_X)(Y - m_Y)\right].$$

Для НВВ:

$$K_{XY} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - m_X)(y - m_Y) f(x,y)dxdy.$$

Для кількісної характеристики залежності ВВ часто використовують (особливо у статистиці) **коефіцієнт кореляції**: $r_{XY} = \dfrac{K_{XY}}{\sigma_X \sigma_Y}$.

Якщо ВВ X та Y дискретні, то у вищенаведених формулах знаки інтегралів замінюють знаками суми по усім можливим значенням ВВ.

Означення. ВВ X та Y називають **некорельованими**, якщо їх кореляційний момент або коефіцієнт кореляції дорівнює нулю.

Властивості коефіцієнта кореляції.

1) $|r_{XY}| \le 1$;

2) якщо X та Y незалежні, то $r_{XY} = 0$;

3) якщо між X та Y є лінійна залежність $Y = aX + b$, де a, b - постійні, то $r_{XY} = 1$.

Зауваження. ***Якщо кореляційний момент або коефіцієнт кореляції відмінний від нуля, то ВВ X та Y - корельовані. Дві корельовані ВВ обов'язково залежні. Але залежні ВВ можуть бути як корельованими, так і некорельованими, тобто їх коефіцієнт кореляції може бути відмінним від нуля або рівним нулю. Із незалежності***

ВВ випливає їх некорельованість, але із некорельованості не випливає незалежність ВВ. У випадку нормально розподілених ВВ із некорельованості випливає незалежність ВВ.

Функції ВВ та їх характеристики.

Означення. Якщо вказано закон (або правило) f, за яким кожному можливому значенню ВВ X відповідає певне значення ВВ Y, то Y називають **функцією** X і позначають $Y = f(X)$.

Відзначимо, що іноді різним можливим значенням ВВ X відповідають однакові значення ВВ Y. Наприклад, якщо $Y = X^2 + 5$, то значенням ± 3 ВВ X відповідає одне значення ВВ $Y = 14$.

Однією із задач теорії імовірностей є визначення законів розподілу та числових характеристик функцій випадкового аргументу, закон розподілу якого відомий.

Нехай $Y = f(X)$, а аргумент X - ДВВ, наприклад, задана таблицею

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

Тоді $Y = f(X)$ також є ДВВ, закон розподілу якої буде мати вигляд:

Y	$f(x_1)$	$f(x_2)$	...	$f(x_n)$
P	p_1	p_2	...	p_n

Математичне сподівання, дисперсію, середнє квадратичне відхилення та початкові і центральні моменти розподілу знаходять за формулами:

$$M(Y) = \sum_{i=1}^{n} f(x_i)\, p_i ;$$

$$D(Y) = M(Y^2) - M^2(Y) = \sum_{i=1}^{n} \left[f(x_i) \right]^2 p_i - M^2(Y);$$

$$\sigma(Y) = \sqrt{D(Y)};$$

$$\nu_k = \sum_{i=1}^{n} \left[f(x_i) \right]^k p_i ;$$

$$\mu_k = \sum_{i=1}^{n} \left[f(x_i) - M(Y) \right]^k p_i .$$

Приклад. ДВВ X задана таблицею

X	-2	0	2
P	$0{,}2$	$0{,}5$	$0{,}3$

Знайти закон розподілу та числові характеристики функції $Y = X^2 + 5$.

Розв'язування.

Нехай X - НВВ, закон розподілу якої заданий диференціальною функцією (щільністю розподілу імовірностей) $f(x)$, а ВВ $Y = \varphi(X)$. Якщо φ - диференційовна функція, монотонна на усьому проміжку можливих значень X, то щільність розподілу функції $Y = \varphi(X)$ визначається за формулою:

$$g(y) = f\left[\psi(y)\right] \cdot \left|\psi'(y)\right|, \qquad (*)$$

де $\psi(y)$ - функція, обернена до функції $\varphi(x)$.

Якщо $\varphi(x)$ немонотонна функція на області визначення аргументу X, то обернена функція неоднозначна, тому щільність розподілу $g(y)$ визначається як сума доданків, кількість яких дорівнює кількості значень оберненої функції:

$$g(y) = \sum_{i=1}^{k} f\left[\psi_i(y)\right] \cdot \left|\psi_i'(y)\right| , \qquad (**)$$

де $\psi_i(y)$ - обернені функції при заданому y.

Алгоритм знаходження щільності розподілу $g(y)$.

1. Визначити множину можливих значень для $Y = \varphi(X)$.

2. Із функціональної залежності $Y = \varphi(X)$ *знайти явний вираз* X *через* Y, *тобто функцію* $\psi(y)$, *обернену до функції* $\varphi(x)$.

3. Знайти похідну $\psi'(y)$.

4. За формулою (*) записати щільність розподілу ВВ Y.

5. Перевірити умову нормування для $g(y)$: $\int_{-\infty}^{+\infty} g(y)dy = 1$.

Приклад. ВВ X розподілена за нормальним законом з математичним сподіванням $a = 0$ та середнім квадратичним відхиленням $\sigma = 1$ (стандартний нормальний розподіл). Знайти закон розподілу функції $Y = X^3$.

Розв'язування.

Для знаходження числових характеристик функції $Y = \varphi(X)$ можна спочатку знайти щільність розподілу $g(y)$ за формулою (*) або (**), а потім скористуватись означеннями. Але інколи можна знайти числові характеристики безпосередньо. Наприклад, математичне сподівання функції $Y = \varphi(X)$ можна знайти за формулою

$$M(\varphi(X)) = \int_{-\infty}^{+\infty} \varphi(x) f(x)dx,$$

де $f(x)$ - щільність розподілу імовірностей ВВ X.

Приклад. НВВ X задана щільністю розподілу імовірностей

$$f(x) = \begin{cases} 1/4 & , \quad x \in [1;5] \\ 0 & , \quad x \notin [1;5] \end{cases}$$

Знайти числові характеристики функції $Y = X^2$.

Розв'язування.

Функції двох неперервних випадкових аргументів

Якщо кожній парі можливих значень ВВ X та Y відповідає одне можливе значення ВВ Z, то Z називають **функцією двох випадкових аргументів** і записують $Z = \varphi(X, Y)$.

Композицією розподілів (згорткою) називається формула, за якою можна отримати закон розподілу **суми незалежних ВВ**.

Зазначимо, що закон розподілу суми не співпадає, взагалі кажучи, із законами розподілу доданків (навіть, якщо вони однаково розподілені).

Нехай НВВ X та Y задані щільностями розподілу відповідно $\varphi_1(x)$ та $\varphi_2(y)$. Тоді щільність розподілу $\varphi(z)$ їх суми $Z = X + Y$ (за умови, що щільність розподілу хоча б одного із аргументів задана на інтервалі $(-\infty; +\infty)$ одною формулою) можна знайти за формулою, яку називають **формулою композиції двох розподілів (формулою згортки)**:

$$\varphi(z) = \int_{-\infty}^{+\infty} \varphi_1(x)\varphi_2(z-x)dx = \int_{-\infty}^{+\infty} \varphi_1(z-y)\varphi_2(y)dy.$$

Зауваження. Якщо можливі значення аргументів невід'ємні, то щільність розподілу $\varphi(z)$ їх суми $Z = X + Y$ знаходять за формулою:

$$\varphi(z) = \int_{0}^{z} \varphi_1(x)\varphi_2(z-x)dx = \int_{0}^{z} \varphi_1(z-y)\varphi_2(y)dy$$

Приклад. Довести, що композиція двох незалежних ВВ, що розподілені за нормальними стандартними законами, також є нормально розподіленою ВВ.

Приклад. Скласти композицію двох незалежних ВВ, що розподілені рівномірно на відрізку [0;1].

Розв'язування.

Розділ 8. Математична статистика

Предмет математичної статистики полягає у розробці методів збору та обробки статистичних даних для одержання наукових та практичних висновків.

Основні задачі математичної статистики:

1) вказати способи збору та групування (якщо даних дуже багато) статистичних відомостей;
2) визначити закон розподілу випадкової величини або системи випадкових величин за статистичними даними;
3) визначити невідомі параметри розподілу;
4) перевірити правдоподібність припущень про закон розподілу випадкової величини, про форму зв'язку між випадковими величинами або про значення параметру, який оцінюють.

ГЕНЕРАЛЬНА ТА ВИБІРКОВА СУКУПНОСТІ.

Нехай потрібно вивчити сукупність об'єктів відносно деякої **якісної або кількісної ознаки (випадкової величини)**, які характеризують ці об'єкти. Кожен об'єкт, який спостерігають, має декілька ознак. Розглядаючи лише одну ознаку кожного об'єкта, ми припускаємо, що інші ознаки рівноправні, або що **множина об'єктів однорідна.**

Такі множини однорідних об'єктів називають статистичною сукупністю.

Наприклад, якщо досліджують партію деталей, то **якісною ознакою** може бути стандартність або нестандартність кожної деталі, а **кількісною ознакою** – розмір деталі. Кількісні ознаки бувають **дискретними та неперервними.**

Перевірку статистичної сукупності можна провести двома способами:

1) перевірити усі об'єкти сукупності (суцільна перевірка або **перепис**);
2) перевірити лише певну частину об'єктів сукупності (**вибірка**).

Перевагами вивчення вибірки є малі затрати коштів, обладнання та часу. Вибірку можна ефективно застосовувати для вивчення відповідної ознаки усієї сукупності лише тоді, коли дані вибірки вірно відображають цю ознаку, тобто вибірка повинна бути **репрезентативною (представницькою, показною).** Згідно із законом великих чисел теорії імовірностей можна стверджувати , що вибірка буде репрезентативною лише тоді, коли її здійснюють випадково.

Простим випадковим відбором (простою випадковою вибіркою) називають такий відбір із статистичної сукупності, при якому кожний об'єкт, що відбирається, має однакову імовірність потрапити до вибірки. **Об'єм n вибіркової сукупності (вибірки) – це кількість об'єктів цієї сукупності.**

Варто відмітити, що альтернативою для простої випадкової вибірки в статистиці є **розшарована випадкова вибірка.**

Генеральною називають сукупність об'єктів, з якої зроблено вибірку. Об'єм генеральної сукупності позначають N .

Вибірки бувають повторні (при яких відібраний об'єкт повертається до генеральної сукупності перед відбором іншого об'єкта) та **безповторні** (при яких взятий об'єкт до генеральної сукупності не повертається). Найчастіше використовуються безповторні вибірки.

ДЖЕРЕЛА ДАНИХ У СТАТИСТИЦІ.

Дослідники і менеджери отримують дані, необхідні для прийняття рішень, в основному, з трьох джерел:

1) вибіркові обстеження;
2) спеціально поставлені експерименти;
3) результати повсякденної (рутинної) роботи у бізнесі.

Приклади: 1) дослідницький центр вибирає 1000 потенційних виборців для опитування з метою вивчення рейтингу певного кандидата на виборах; 2) проведення анкетування серед певної групи людей за спеціально розробленою анкетою; 3) аналіз даних рівня продажу певного виду товару, різноманітні офіційні джерела статистичних даних.

Джерела даних бувають **первинними** та **вторинними**.

Первинні дані збираються спеціально для статистичного дослідження. Для цих даних є відомості про методи збирання, точність даних тощо.

Вторинними є дані, що використовуються у статистиці, але спочатку збирались для інших цілей. Очевидно, що рутинні записи про діяльність фірм, офіційні статистичні звіти є вторинними даними.

Безумовно, більш цінними є первинні дані, але їх не завжди можна отримати, тому часто використовуються вторинні дані.

СПОСОБИ ВІДБОРУ.

1. Вибір, який не потребує розділення генеральної сукупності на частини. До цього вибору відносять:
   - простий випадковий безповторний відбір;
   - простий випадковий повторний відбір.
2. Вибір, при якому генеральна сукупність розділяється на частини (розшарований випадковий відбір). До цього виду вибору відносять:
   - типовий відбір;
   - механічний відбір;
   - серійний відбір.

Типовим називають відбір, при якому об’єкти відбирають не із усієї генеральної сукупності, а лише із її типових частин. Наприклад, якщо однакові вироби виготовляються на різних підприємствах (або різними станками), то відбираються вироби кожного окремого підприємства (станка) тощо.

Механічним називають відбір, при якому генеральна сукупність механічно поділяється на стільки частин, скількимає бути об’єктів у вибірці. Із кожної частини випадковим чином відбирають один об’єкт. Наприклад, якщо потрібно перевірити 25% усіх виготовлених виробів, то відбирають кожний четвертий виріб. Зазначимо, що для репрезентативності механічного відбору потрібно враховувати специфіку технологічного процесу.

Серійним називають відбір, при якому об’єкти із генеральної сукупності відбирають не по одному, а серіями. Серійний відбір застосовують тоді, коли ознака, яку досліджують, мало змінюється у різних серіях.

Зауважимо, що в економічних дослідженнях застосовують і **комбіновані** відбори.

ПРОСТА ВИПАДКОВА ВИБІРКА.

Для здійснення простої випадкової вибірки необхідна наявність **основи вибірки,** тобто такого представлення генеральної сукупності, при якому її елементи були б принаймні перераховані.

Приклад. а) генеральна сукупність – всі клієнти банку. Основою вибірки можуть бути робочі списки клієнтів, що вів банк.

б) генеральна сукупність – всі мешканці міста, які мають стаціонарні телефони. Основою вибірки може бути телефонний довідник.

Як правило, дані для утворення простої випадкової вибірки подаються у вигляді деякої, заздалегідь складеної таблиці і тому основою вибірки є нумерація елементів цієї таблиці.

Основа вибірки повинна повністю відбивати ознаку генеральної сукупності, яка вивчається. Порушення цієї вимоги може зробити вибірку нерепрезентативною.

Приклад. Для обстеження молодих сімей міста на предмет наявності в них дітей дошкільного віку дослідник випадковим чином за допомогою телефонного довідника додзвонюється до сім’ї з 18.00 до 21.00 щодня. Чи буде така вибірка репрезентативною?

Приклад. Проста випадкова вибірка може використовуватись у наступних дослідженнях:

а) телефонна компанія перевіряє рахунки 10% всіх міжнародних телефонних переговорів з метою визначення їх середньої величини;

б) аудиторська перевірка 20% фірм регіону з метою контроля правильності сплати податків.

Загальновідомо, що найкращим способом здійснення простої випадкової вибірки є використання випадкових вибіркових чисел (їх таблиць або за допомогою стандартних комп'ютерних програм, зокрема, функції "выборка" електронних таблиць Excel).

СТАТИСТИЧНИЙ РОЗПОДІЛ ОЗНАКИ.

Дані у статистиці, отримані за допомогою спеціальних досліджень або із робочих (рутинних) записів у бізнесі, надходять до дослідника у вигляді неорганізованої маси (незалежно від того, чи є вони вибірковими, чи даними із генеральної сукупності). В математичній статистиці замість слова "**дані**" вживається термін "**варіанти**". Характеристику варіанти (**випадкову величину**) при цьому називають **ознакою.**

Нехай із генеральної сукупності взята вибірка об'єктів $\{x_1, x_2, ..., x_n\}$ об'єму n, для вивчення ознаки X. Тобто, значення $x_1, x_2, ..., x_n$ є **варіанти ознаки** X. Першим кроком обробки є впорядкування варіант. Розглянемо приклад:

Вибірка середньомісячної зарплати 100 співробітників фірми									
338	348	304	314	326	314	324	304	342	308
336	304	302	338	314	304	320	321	322	321
312	323	336	324	312	312	364	356	362	302
322	310	334	292	362	381	304	366	298	304
381	368	304	298	368	290	340	328	316	322
302	314	292	342	321	322	290	332	298	296
296	298	324	338	352	326	318	304	332	322
360	312	331	331	304	316	332	282	342	338
342	322	324	325	302	328	354	330	316	324
334	350	334	324	332	340	324	314	326	323

Розташуємо дані у порядку зростання:

Впорядкована вибірка середньомісячної зарплати 100 співробітників фірми (у порядку зростання)									
282	298	304	314	321	323	326	332	340	356
290	302	304	314	321	324	326	334	340	360
290	302	304	314	321	324	328	334	342	362
292	302	304	314	322	324	328	334	342	362
292	302	308	314	322	324	330	336	342	364
296	304	310	316	322	324	331	336	342	366
296	304	312	316	322	324	331	338	348	368
298	304	312	316	322	324	332	238	350	368
298	304	312	318	322	325	332	338	352	381
298	304	312	320	323	326	332	338	354	381

Варіанти, записані до таблиці у зростаючому (спадаючому) порядку, називають **варіаційним рядом**. При упорядкуванні (**ранжуванні**) можна отримати більше інформації, наприклад, про межі зміни середньомісячної зарплати.

РОЗПОДІЛ ЧАСТОТ.

Нехай у вибірці із n варіант $x_1, x_2, ..., x_n$ ознака X прийняла значення x_1 – m_1 раз, значення x_2 – m_2 раз, …, значення $x_k - m_k$ раз.

Додатне число, що вказує, скільки раз та чи інша варіанта зустрічається в таблиці даних, називається **частотою**, а ряд $m_1, m_2, ..., m_k$ називається **рядом частот**. Відмітимо, що сума усіх частот повинна дорівнювати об’єму вибірки: $\sum_{i=1}^{k} m_i = n$.

Статистичний розподіл вибірки встановлює зв’язок між рядом варіант, що зростає або спадає, і відповідними частотами. Як правило, його подають у вигляді таблиці:

$X = x_i$	x_1	x_2	…	x_k
$m = m_i$	m_1	m_2	…	m_k

Заданий такою таблицею розподіл називають **простим незгрупованим статистичним розподілом або розподілом частоти варіанти (рядом розподілу частоти варіанти).**

Розподіл частоти середньомісячної зарплати співробітників фірми							
$X = x_i$	m_i	$X = x_i$	m_i	$X = x_i$	m_i	$X = x_i$	m_i
282	1	314	5	328	2	350	1
290	2	316	3	330	1	352	1
292	2	318	1	331	2	354	1
296	2	320	1	332	4	356	1
298	4	321	3	334	3	360	1
302	4	322	6	336	2	362	2
304	9	323	2	338	4	364	1
308	1	324	7	340	2	366	1
310	1	325	1	342	4	368	2
312	4	326	3	348	1	381	2

$$\sum_{i=1}^{k=40} m_i = n = 100.$$

Подальшим кроком в обробці даних, що призводить до спрощення досліджень, є їх **згрупування.** Як видно із останньої таблиці максимальне та мінімальне значення варіанти будуть

$$x_{\max} = 381, \quad x_{\min} = 282.$$

Різниця цих чисел

$$R = x_{\max} - x_{\min} = 381 - 282 = 99$$

називається **варіаційним розмахом або розмахом варіант.**

Введемо для варіанти інтервали зміни середньої зарплати: 280-290, 290-300,…, 380-390. Кожний інтервал називається **класом інтервалів або класом**, а число одиниць виміру у цих класах, тобто різниця $h = x_{i+1} - x_i$, називається **шириною класу**. Використовуючи дані попередньої таблиці, отримуємо:

Згрупований розподіл частоти середньомісячної зарплати співробітників фірми	
x	m_i
280-290	1
290-300	10
300-310	14
310-320	14
320-330	25
330-340	16
340-350	7
350-360	4
360-370	7
370-380	0
380-390	2
сума	100

Така таблиця, яка встановлює зв'язок між згрупованим рядом варіант, що зростає або спадає, та сумами їхніх частот по класах, називається **згрупованим розподілом частоти варіанти**. У нашому прикладі ширина класів однакова і дорівнює $h = 10$, а кількість класів $k = 11$. Зазначимо, що введені величини варіаційний розмах, ширина класів та їх кількість пов'язані співвідношенням

$$h = \frac{R}{k}.$$

Зауваження. Інколи неможливо або небажано вибирати ширину класів однаковою. Неоднакова ширина класів бажана, наприклад, коли значення частоти одного чи декількох класів набагато більша (менша) значень частот інших інтервалів. Як правило, ширина інтервалів зростає (або спадає) і може містити інтервали відкритого типу "більше ніж…", "менше ніж…".

ЗГРУПОВАНИЙ РОЗПОДІЛ НАКОПИЧЕНОЇ ЧАСТОТИ.

Часто поряд із розподілом частоти варіанти необхідно мати розподіл **пакопиченої (кумулятивної) частоти**. Такий розподіл одержується послідовним додаванням частот чергового інтервалу, починаючи з першого і зікінчуючи останнім (див.таблицю):

Згрупований розподіл частоти середньомісячної зарплати співробітників фірми			
інтервали платні X	частоти m_i	платня	накопичені частоти F_i
280-290	1	<290	1
290-300	10	<300	11
300-310	14	<310	25
310-320	14	<320	39
320-330	25	<330	64
330-340	16	<340	80
340-350	7	<350	87
350-360	4	<360	91
360-370	7	<370	98
370-380	0	<380	98
380-390	2	<390	100
сума	100		

Розподіл накопиченої частоти дозволяє відповісти на питання: “Скільки існує варіант, менших, наприклад, 350?” Із таблиці знаходимо: $F_{350} = 87$.

РОЗПОДІЛ ЧАСТКИ (ВІДНОСНОЇ ЧАСТОТИ).

Часто замість значень частот використовуються відношення частоти m_i варіанти x_i до об’єму вибірки n :

$$w_i = \frac{m_i}{n},$$

які називаються **частками (відносними частотами)**, причому $\sum_{i=1}^{k} w_i = 1$.

Залежність між впорядкованим рядом варіант і відповідними їм частками також називають **статистичним розподілом вибірки** (див. таблицю):

Згрупований розподіл частки w_i та накопиченої частки F_i / n середньомісячної зарплати співробітників фірми					
інтервали платні x	частоти m_i	частки w_i	платня	накопичені частоти F_i	накопичені частки F_i / n
280-290	1	0,01	<290	1	0,01
290-300	10	0,1	<300	11	0,11
300-310	14	0,14	<310	25	0,25
310-320	14	0,14	<320	39	0,39
320-330	25	0,25	<330	64	0,64
330-340	16	0,16	<340	80	0,80
340-350	7	0,07	<350	87	0,87
350-360	4	0,04	<360	91	0,91
360-370	7	0,07	<370	98	0,98
370-380	0	0,00	<380	98	0,98
380-390	2	0,02	<390	100	1
сума	100	1			

Розподіл накопиченої частки дозволяє відповісти на питання: “Яка частка варіант, що менші, наприклад, 350?” Із таблиці знаходимо: частка цих варіант становить 0,87.

ЗГРУПОВАНИЙ РОЗПОДІЛ ЩІЛЬНОСТЕЙ ЧАСТОТИ ТА ЧАСТКИ.

Якщо поділити всі частоти на ширину інтервалу, то отримаємо **розподіл щільності частоти** вибірки: $\frac{m_i}{h}$.

Відзначимо, що поняття щільностей мають глибокий імовірністний смисл.

Уведемо до попередньої таблиці стовпці щільностей частот та часток:

інтервали платні X	часто-ти m_i	частки w_i	щіль-ність часто-ти $\frac{m_i}{h}$	щіль-ність частки $\frac{w_i}{h}$	платня	накопичені частоти F_i	накопичені частки F_i / n
280-290	1	0,01	0,1	0,001	<290	1	0,01
290-300	10	0,1	1,0	0,010	<300	11	0,11
300-310	14	0,14	1,4	0,014	<310	25	0,25
310-320	14	0,14	1,4	0,014	<320	39	0,39
320-330	25	0,25	2,5	0,025	<330	64	0,64
330-340	16	0,16	1,6	0,016	<340	80	0,80
340-350	7	0,07	0,7	0,007	<350	87	0,87
350-360	4	0,04	0,4	0,004	<360	91	0,91
360-370	7	0,07	0,7	0,007	<370	98	0,98
370-380	0	0,00	0,0	0,000	<380	98	0,98
380-390	2	0,02	0,2	0,002	<390	100	1
сума	100	1					

ЗАГАЛЬНА СХЕМА ПОБУДОВИ ЗГРУПОВАНОГО РОЗПОДІЛУ ЧАСТОТ.

1. Визначити найбільше $x_{\max}$ та найменше $x_{\min}$ значення варіанти x_i і визначити варіаційний розмах $R = x_{\max} - x_{\min}$.

2. Задатися певним числом класів k, яке рекомендується брати непарним, при об'ємах вибірки $n \geq 100$ доцільно $9 \leq k \leq 15$, а при менших об'ємах вибірки можна $k = 7$.

3. Визначити ширину класів $h = R / k$. Для спрощення розрахунків, отримане значення ширини класів слід округлити до найближчого цілого.

4. Встановити границі класів і підрахувати кількість варіант у кожному класі.
5. Визначити частоту для кожного класу і записати ряд розподілу.

ЕМПІРИЧНА ФУНКЦІЯ РОЗПОДІЛУ.

Нехай є статистичний розподіл частот деякої ознаки X.

Означення. **Емпіричною функцією розподілу (або функцією розподілу вибірки)** називають функцію $F^e(x)$, яка визначає для кожного дійсного значення x частку події $X < x$, тобто:

$$F^e(x) = \frac{m_x}{n},$$

де m_x - частота (кількість) варіант, які менші від x, а n - об'єм вибірки.

Зауваження. Інтегральну функцію розподілу $F(x)$ генеральної сукупності в математичній статистиці називають **теоретичною функцією розподілу**. Вона відрізняється від емпіричної функції розподілу $F^e(x)$ тим, що визначає імовірність події $X < x$, а не її частку. Із теореми Бернуллі випливає, що частка $F^e(x) = \frac{m_x}{n}$ події $X < x$ прямує до імовірності $F(x) = P(X < x)$ цієї події. Тому доцільно використовувати емпіричну (вибіркову) функцію розподілу в якості наближення до теоретичної функції розподілу генеральної сукупності.

Між емпіричною функцією розподілу $F^e(x)$ і функцією накопичених частот на кожному класі інтервалів F_i існує простий зв'язок:

$$F^e(x) = \frac{m_x}{n} = \frac{F_i}{n}.$$

ГРАФІЧНЕ ЗОБРАЖЕННЯ СТАТИСТИЧНИХ РОЗПОДІЛІВ.

Полігоном частот називають ламану, відрізки якої з'єднують точки $(x_1; m_1), (x_2; m_2), \ldots, (x_k; m_k)$.

Полігоном часток (відносних частот) називають ламану, відрізки якої з'єднують точки $(x_1; w_1),(x_2; w_2),...,(x_k; w_k)$.

Полігон часток є аналогом полігону розподілу імовірностей.

Ці полігони використовують для графічного зображення дискретних варіаційних рядів. А для зображення інтервальних варіаційних рядів використовують гістограми.

Гістограмою частот називають ступінчасту фігуру, що складається з прямокутників, основами яких є частинні інтервали варіант довжиною $h = x_{i+1} - x_i$, а висоти дорівнюють $\frac{m_i}{h}$ (**щільність частоти**). **Площа гістограми частот дорівнює об'єму вибірки.**

Гістограмою часток (відносних частот) називають ступінчасту фігуру, що складається з прямокутників, основами яких є частинні інтервали варіант довжиною $h = x_{i+1} - x_i$, а висоти дорівнюють $\frac{w_i}{h}$ (**щільність частки**). **Площа гістограми частки дорівнює 1. Гістограма частки є аналогом розподілу щільності імовірностей для генеральної сукупності.**

Полігони накопиченої частки (або накопиченої частоти) в статистиці називають **огівою або кумулятивною кривою.**

Графіки статистичних розподілів для розглянутого прикладу побудовані на окремих листах засобами Excel (див. додатки).

ОСНОВНІ ВИМОГИ ДО СТАТИСТИЧНИХ ОЦІНОК ПАРАМЕТРІВ РОЗПОДІЛУ.

У багатьох випадках потрібно дослідити кількісну ознаку X генеральної сукупності, використовуючи результати вибірки. Часто для цього достатньо знати наближені значення математичного сподівання $M(X)$, дисперсії $D(X)$, середньоквадратичного відхилення $\sigma(X)$, початкові або центральні моменти. Іноді з деяких міркувань вдається встановити закон розподілу X. Тоді треба вміти оцінювати параметри цього закону розподілу.

Означення. **Статистичною (точковою) оцінкою** невідомого параметра випадкової величини X генеральної сукупності (теоретичного розподілу X) називають функцію від випадкових величин (результатів вибірки), що спостерігаються.

Нехай θ^* є статистична оцінка невідомого параметра θ теоретичного розподілу. Припустимо, що за вибіркою об'єму n знайдена оцінка θ_1^*. При інших вибірках того ж об'єму одержимо деякі інші оцінки $\theta_2^*, \theta_3^*, ..., \theta_k^*$. Саме оцінку θ^* можна розглядати як випадкову величину, а числа $\theta_1^*, \theta_2^*, ..., \theta_k^*$ як її можливі значення. Точкова статистична оцінка повинна задовольняти певним умовам, які сформулюємо у вигляді означень.

Означення. Статистичну оцінку θ^* параметра θ називають **незсунутою**, якщо $M(\theta^*) = \theta$. Оцінку θ^* називають **зсунутою**, якщо ця рівність не виконується.

Означення. Статистичну оцінку θ^* параметра θ називають **ефективною**, якщо вона при заданому об'ємі вибірки n має найменшу можливу дисперсію.

Означення. Статистичну оцінку θ^* параметра θ називають **обгрунтованою (значимою, показною, репрезентативною)**, якщо вона при $n \to \infty$ прямує за імовірністю до оцінюваного параметра.

Відмітимо, що якщо дисперсія незсунутої оцінки при $n \to \infty$ прямує до нуля, то оцінка буде і обгрунтованою.

ЧИСЛОВІ ХАРАКТЕРИСТИКИ ВИБІРКИ.

Окрім табличних та графічних методів представлення даних широко застосовуються їх числові характеристики. Найбільш важливі із них: **середнє значення, дисперсія, середнє квадратичне відхилення (стандартне відхилення).** Ці характеристики називають **генеральними**, якщо вони обчислені за даними генеральної сукупності, та **вибірковими**, якщо вони обчислені за даними вибірки.

Числові характеристики, обчислені по вибірці або ті, що використовуються для опису даних вибірки, часто називають **статистиками**.

Числові характеристики, обчислені по генеральній сукупності або ті, що використовуються для опису даних генеральної сукупності, часто називають **параметрами**.

По аналогії із математичним сподіванням, дисперсією та середнім квадратичним відхиленням ДВВ обчислюються вибіркові характеристики (статистики), замінюючи при цьому відповідні імовірності p_i частками $\frac{m_i}{n}$, що відповідають варіантам x_i (якщо неперервна ознака задана інтервальним варіаційним рядом, то проводять його "дискретизацію", замінюючи кожний інтервал його середнім значенням).

Означення. **Вибірковою середньою або вибірковою зваженою середньоарифметичною** називають середню арифметичну варіант вибірки із урахуванням їх частостей і позначають

$$X_B = \tilde{X} = \sum_{i=1}^{k} x_i \frac{m_i}{n} = \frac{1}{n} \sum_{i=1}^{k} x_i m_i \ ,$$

де n - об'єм вибірки, k - кількість різних варіант, m_i - частоти варіант x_i ($m_1 + ... + m_k = n$). Аналогічно визначається **генеральна середня або генеральна зважена середньоарифметична** із заміною об'єму вибірки n на N - об'єм генеральної сукупності і позначається $X_Г = \bar{X}$.

Вибіркова середня є аналогом математичного сподівання і використовується дуже часто. Вона може приймати різні числові значення при різних вибірках однакового об'єму. Тому можна розглядати розподіли вибіркової середньої та числові характеристики цього розподілу. Неважко довести, що:

Теорема. Вибіркова середня є незсунутою, ефективною та обгрунтованою точковою оцінкою для генеральної середньої. Іншими словами, вибіркова середня є статистикою, яка задовольняє всі умови точкового оцінювання, для параметра – генеральної середньої кількісної ознаки.

Зауваження. Крім вибіркової середньої (середньозваженої) в статистиці застосовуються і інші середні, зокрема: **проста середньоарифметична** $\bar{X} = \frac{1}{n} \sum_{i=1}^{k} x_i$; **степеневі середні (середньоквадратична, середня гармонічна, середня геометрична тощо)** ; **структурні середні, які не залежать від значень варіант, що розташовані на краях розподілу, зокрема, мода** Mo**(значення варіанти, яка має найбільшу частоту) та медіана** Me**(значення, яке "ділить розподіл навпіл") та інші.**

Означення. **Вибірковою дисперсією** називають середню (зважену) квадратів відхилення варіант від вибіркової середньої:

$$D(\tilde{X}) = D_B(X) = \sigma_B^2 = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \tilde{X}\right)^2 m_i \ .$$

Зауважимо, що для спрощення обчислення вибіркової дисперсії можна застосовувати формулу: $D_B(X) = \tilde{X}^2 - \left(\tilde{X}\right)^2$.

Означення. **Вибірковим середньоквадратичним відхиленням (стандартом)** називають квадратний корінь із вибіркової дисперсії $\sigma_B = \sqrt{D_B}$.

Можна показати, що:

Вибіркова дисперсія D_B є ефективною, обгрунтованою, але ЗСУНУТОЮ точковою оцінкою для генеральної дисперсії $D_o = D(\bar{X}) = D_\Gamma = \sigma_\Gamma^2$.

Зауваження. Вибіркова дисперсія дає занижені оцінки для генеральної дисперсії, але $M(D_B) = \frac{n-1}{n} D_o$. Тому вибіркову дисперсію виправляють так, щоб вона стала незсунутою оцінкою. А саме, вводять так звану **виправлену вибіркову дисперсію** $s^2 = \frac{n}{n-1} D_B$.

Тоді **виправленим стандартом вибірки** буде $s = \sqrt{s^2}$.

Очевидно, що при достатньо великих об'ємах вибірки ($n \geq 30$) вибіркова дисперсія та виправлена вибіркова дисперсія різняться дуже мало, тому в практичних задачах виправлені вибіркові дисперсію та стандарт використовують лише при об'ємах вибірок $n < 30$.

Окрім вищевказаних числових характеристик, використовують статистичні початкові та центральні моменти інших порядків, зокрема, коефіцієнт асиметрії As, який характеризує "скошеність" розподілу відносно його центра та коефіцієнт ексцесу Es, який характеризує "крутизну" розподілу.

У випадку оцінювання параметрів якісної ознаки X, а саме генеральної частки (частості) $p=\frac{m}{N}$ за допомогою вибіркової частки $w=\frac{m}{n}$, використовується наступна

Теорема. Вибіркова частка w є незсунутою, ефективною та обгрунтованою точковою оцінкою для генеральної частки p . Іншими словами, вибіркова частість w є статистикою, яка задовольняє всі умови точкового оцінювання, для параметра – генеральної частості p якісної ознаки.

ІНТЕРВАЛЬНІ ОЦІНКИ.

Точкові оцінки параметрів розподілу є випадковими величинами, їх можна вважати первинними результатами обробки вибірки, оскільки невідомо, з якою точністю кожна з них оцінює відповідну числову характеристику генеральної сукупності. Якщо об'єм вибірки досить великий, то точкові оцінки задовольняють практичні потреби точності. Якщо ж об'єм вибірки малий, то точкові оцінки можуть давати значні похибки, тому питання точності оцінювання у цьому випадку дуже важливе і необхідно використовувати інтервальні оцінки.

Означення. **Інтервальною** називають оцінку, яка визначається двома числами – кінцями інтервалу. Інтервальні оцінки дозволяють встановити точність та надійність оцінок.

Нехай знайдена за даними вибірки статистична оцінка θ^* є точковою оцінкою невідомого параметра θ. Очевидно, що θ^* тим точніше визначає θ , чим меншим є модуль різниці $\theta^*-\theta$. Іншими словами, якщо $\left|\theta-\theta^*\right|\le\Delta \quad , \quad \Delta>0$, тоді меншому Δ відповідатиме більш точна оцінка. Тому число Δ називають **граничною похибкою вибірки** і воно характеризує точність оцінки.

Але статистичні методи не дозволяють категорично стверджувати, що оцінка θ^* задовольняє нерівність $\left|\theta-\theta^*\right|\le\Delta$. Таке твердження можна зробити лише із певною імовірністю.

Означення. **Надійністю (довірчою імовірністю)** оцінки θ^* параметра θ називають імовірність

$$\gamma = P = P\left(\left|\theta - \theta^*\right| \le \Delta\right),$$

яку можна записати у вигляді $\gamma = P\left(\theta^* - \Delta \le \theta \le \theta^* + \Delta\right)$. З цієї рівності випливає, що інтервал $\left(\theta^* - \Delta; \theta^* + \Delta\right)$ містить невідомий параметр θ генеральної сукупності (часто кажуть, що інтервал покриває невідомий параметр).

Означення. Інтервал $\left(\theta^* - \Delta; \theta^* + \Delta\right)$ називають **довірчим**, якщо він покриває невідомий параметр θ із заданою надійністю γ.

Зауважимо, що кінці довірчого інтервалу є випадковими величинами.

За допомогою теорем закону великих чисел з уточненням Ляпунова (Чебишова для кількісної ознаки та Бернуллі для якісної ознаки) доводиться наступне твердження (**класичні інтервальні оцінки або формули довірчої імовірності**):

Теорема. Імовірність того, що модуль відхилення вібіркової середньої (або частки) від генеральної середньої (або частки) не перевищить число $\Delta > 0$ дорівнює:

$\gamma = P\left(\left|X_B - X_\Gamma\right| \le \Delta\right) = 2\Phi(t)$ (або $\gamma = P\left(\left|w - p\right| \le \Delta\right) = 2\Phi(t)$),

де $\Phi(t)$ - інтегральна функція Лапласа, $t = \dfrac{\Delta}{\mu}$, а μ - **середньоквадратична похибка (стандарт) вибірки**, яка може бути знайдена за наступними формулами:

а) при оцінюванні середньої кількісної ознаки:

$$\mu = \sqrt{\frac{\sigma_\Gamma^2}{n}} \approx \sqrt{\frac{\sigma_B^2}{n}}$$ у випадку повторної вибірки,

$$\mu = \sqrt{\frac{\sigma_\Gamma^2}{n}\left(1 - \frac{n}{N}\right)} \approx \sqrt{\frac{\sigma_B^2}{n}\left(1 - \frac{n}{N}\right)}$$ у випадку безповторної вибірки;

б) при оцінюванні частки якісної ознаки

$$\mu = \sqrt{\frac{pq}{n}} \approx \sqrt{\frac{w(1-w)}{n}}$$ у випадку повторної вибірки,

$$\mu = \sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)} \approx \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$$ у випадку безповторної вибірки.

Зауваження. При визначенні середньоквадратичної похибки вибірки для частки якісної ознаки буває, що невідомі ні генеральна частка p, ні її точкова оцінка – вибіркова частка w. Тоді добуток pq покладають рівним максимальному можливому значенню - $0{,}25$.

Наслідок. При заданій надійності (довірчій імовірності) $\gamma = 2\Phi(t)$ гранична похибка вибірки дорівнює t-кратній величині стандарту, тобто

$$\Delta = t\mu.$$

Наслідок. Довірчі інтервали (інтервальні оцінки) для генеральної середньої та генеральної частки визначаються формулами:

$$X_B - \Delta \le X_\Gamma \le X_B + \Delta$$

та

$$w - \Delta \le p \le w + \Delta.$$

Із класичних оцінок, в яких точність оцінки визначається граничною похибкою $\Delta = t\mu = \frac{t\sigma_B}{\sqrt{n}}$, можна зробити наступні висновки:

1) при зростанні об’єму вибірки n гранична похибка Δ зменшується, тому точність оцінки збільшується (оскільки звужується довірчий інтервал);
2) збільшення надійності $\gamma = 2\Phi(t)$ оцінки призводить до зростання t (оскільки $\Phi(t)$ - зростаюча функція), а внаслідок цього, і до збільшення Δ. Іншими словами: збільшення надійності класичної оцінки призводить до зменшення її точності;
3) аналіз формул для обчислення середньоквадратичної похибки вибірки μ дозволяє при об’ємах генеральної сукупності $N \to \infty$, або у випадках, коли $N >> n$ (об’єм генеральної сукупності набагато більший об’єму вибірки), застосовувати більш прості формули повторної вибірки.

ТРИ ТИПИ ЗАДАЧ ВИБІРКОВОГО МЕТОДА.

1) Для заданих об'ємові вибірки та довірчому інтервалі знайти надійність оцінки (дано: n і Δ; визначається $\gamma = 2\Phi(t) = P$).
2) При заданих об'ємові вибірки та надійності оцінки знайти довірчий інтервал (дано: n і $\gamma = 2\Phi(t) = P$; визначається Δ та $X_B - \Delta \le X_Г \le X_B + \Delta$ або $w - \Delta \le p \le w + \Delta$).
3) При заданих надійності оцінки та довірчому інтервалі знайти необхідний об'єм вибірки (дано: $\gamma = 2\Phi(t) = P$ і Δ; знаходиться n).

Приклад. За умовами результатів вибірки зарплати 100 співробітників фірми із її 1000 робітників визначити: а) імовірність того, що середня платня всіх робітників фірми відрізняється від середньої вибіркової платні не більше ніж на 5грн. в ту чи іншу сторону; б) границі, в яких з надійністю 0,9545 знаходиться середня платня всіх робітників фірми; в) об'єм вибірки, при якому з надійністю 0,9973 модуль відхилення середньої платні усіх робітників від вибіркової середньої платні не перевищить 5грн. Розглянути випадки повторної та безповторної вибірки.

Розв'язування. Дано: ознака X - платня (кількісна ознака).

Генеральна сукупність:

$N = 1000$ - кількість усіх робітників фірми (об'єм генеральної сукупності);

$X_Г = \bar{X}$ - середня платня усіх робітників (генеральна середня, яка оцінюється).

Вибірка:

$n = 100$ - кількість відібраних робітників (об'єм вибірки);

$X_B = \tilde{X} = 325,8$ вибіркова середня платня,

вибіркові дисперсія $D_B = 513,86$ та стандарт $\sigma_B = 22,67$ (знайдені за умовами попереднього прикладу).

а) Знаходимо середньоквадратичну похибку вибірки. Для повторної вибірки

$\mu \approx \sqrt{\frac{D_B}{n}} = \sqrt{\frac{513,86}{100}} = 2,267$. Довірча імовірність

$P\left(\left|X_Г - X_B\right| \le 5\right) = 2\Phi\left(\frac{5}{2,267}\right) = 2\Phi(2,21) = 0,9722$. Для

безповторної вибірки $\mu \approx \sqrt{\frac{D_B}{n}\left(1-\frac{n}{N}\right)} = \sqrt{\frac{513,86}{100}\left(1-\frac{100}{1000}\right)} = 2,15$, а надійність $P\left(\left|X_{\tilde{A}} - X_B\right| \le 5\right) = 2\Phi\left(\frac{5}{2,15}\right) = 2\Phi(2,33) = 0,9802$.

Отже, імовірність того, що вибіркова середньомісячна платня $X_B = 325,8$ буде відрізнятися по модулю від середньомісячної платні всіх робітників фірми $X_Г$ не більше, ніж на 5грн., буде дорівнювати 0,9722 для повторної і 0,9802 для безповторної вибірки.

б) Знаходимо граничні похибки повторної та безповторної вибірки за формулою $\Delta = t\mu$, в якій $t = 2$ (знаходиться як аргумент значення інтегральної функції Лапласа $\gamma = 2\Phi(t) = 0,9545 \Rightarrow \quad \Phi(t) = 0,47725$ по таблиці або за допомогою спеціальної функції Excel).

Для повторної вибірки $\Delta = 2 \cdot 2,267 = 4,534$, а довірчий інтервал: $325,8 - 4,534 \le X_{\tilde{A}} \le 325,8 + 4,534$ або $321,3 \le \tilde{O}_{\tilde{A}} \le 330,3$ або $\tilde{O}_{\tilde{A}} = 325,8 \pm 4,5$ грн.

Для безповторної вибірки $\Delta = 2 \cdot 2,15 = 4,3$, а довірчий інтервал: $325,8 - 4,3 \le X_{\tilde{A}} \le 325,8 + 4,3$ або $321,5 \le \tilde{O}_{\tilde{A}} \le 330,1$ або $\tilde{O}_{\tilde{A}} = 325,8 \pm 4,3$ грн..

Отже, з надійністю 0,9545 можна стверджувати, що середньомісячна платня всіх робітників фірми буде від 321,3грн. до 330,3грн. у випадку повторної вибірки та від 321,5грн. до 330,1грн. у випадку безповторної вибірки.

в) Знаходимо аргумент значення інтегральної функції Лапласа $\gamma = 2\Phi(t) = 0,9973$ по таблиці або за допомогою спеціальної функції Excel: $t = 3$. У формулу $\Delta = t\mu$ при $\Delta = 5$ підставимо усі відомі величини: $5 = 3\sqrt{\frac{513,86}{n}}$, звідси $n = 185$ для повторної вибірки; $5 = 3\sqrt{\frac{513,86}{n}\left(1-\frac{n}{1000}\right)}$, звідси $n = 157$ для безповторної вибірки.

Приклад. Із партії в 8000 телевізорів відібрано 800. Серед відібраних виявилось 10% нестандартних. Знайти: а) імовірність того, що частка стандартних телевізорів в усій партії відрізняється по модулю від отриманої частки таких телевізорів у вибірці не більш, ніж на 0,02; б) границі, в яких з імовірністю 0,95 знаходиться частка стандартних телевізорів в усій партії; в) кількість телевізорів, які потрібно відібрати, щоб з надійністю 0,9545 частка стандартних телевізорів серед відібраних відрізнялась від генеральної частки по модулю не більш, ніж на 0,03; г) як змінилися б результати попередніх пунктів, якби про частку нестандартних телевізорів взагалі не було б нічого відомо. Розглянути випадки повторної та безповторної вибірки.

Розв'язування.

Дано: ознака X - телевізор стандартний (якісна ознака).

Генеральна сукупність:

$N = 8000$ - об'єм усієї партії;

p - частка стандартних телевізорів в усій партії (оцінюється).

Вибірка: $n = 800$ - об'єм;

$w = \frac{m}{n} = \frac{90\%}{100\%} = 0{,}9$ вибіркова частка стандартних телевізорів.

а) Знаходимо середньоквадратичну похибку вибірки для частки. Для повторної вибірки $\mu \approx \sqrt{\frac{w(1-w)}{n}} = \sqrt{\frac{0{,}9 \cdot 0{,}1}{800}} = 0{,}0106$. Довірча імовірність $P\left(\left|0{,}9 - p\right| \le 0{,}02\right) = 2\Phi\left(\frac{0{,}02}{0{,}0106}\right) = 2\Phi(1{,}89) = 0{,}9412$. Для безповторної вибірки $\mu \approx \sqrt{\frac{w(1-w)}{n}\left(1 - \frac{n}{N}\right)} = \sqrt{\frac{0{,}9 \cdot 0{,}1}{800}\left(1 - \frac{800}{8000}\right)} = 0{,}0101$, а надійність $P\left(\left|0{,}9 - p\right| \le 0{,}02\right) = 2\Phi\left(\frac{0{,}02}{0{,}0101}\right) = 2\Phi(1{,}98) = 0{,}9523$.

Отже, імовірність того, що вибіркова частка стандартних телевізорів $w = 0,9$ буде відрізнятися по модулю від генеральної частки не більше, ніж на 0,02, буде дорівнювати 0,9412 для повторної і 0,9523 для безповторної вибірки.

б) Знаходимо граничні похибки повторної та безповторної вибірки за формулою $\Delta = t\mu$, в якій $t = 1,96$ (знаходиться як аргумент значення інтегральної функції Лапласа $\gamma = 2\Phi(t) = 0,95$ по таблиці або за допомогою спеціальної функції Excel).

Для повторної вибірки $\Delta = 1,96 \cdot 0,0106 = 0,0208$, а довірчий інтервал: $0,9 - 0,0208 \le p \le 0,9 + 0,0208$ або $0,8792 \le p \le 0,9208$ або $p = 0,9 \pm 0,0208$ або $p = 90\% \pm 2,08\%$.

Для безповторної вибірки $\Delta = 1,96 \cdot 0,0101 = 0,0198$, а довірчий інтервал: $0,9 - 0,0198 \le p \le 0,9 + 0,0198$ або $0,8802 \le p \le 0,9198$ або $p = 0,9 \pm 0,0198$ або $p = 90\% \pm 1,98\%$.

Отже, з надійністю 0,95 можна стверджувати, що частка стандартних телевізорів в усій партії буде від 0,8792 до 0,9208 у випадку повторної вибірки та від 0,8802 до 0,9198 у випадку безповторної вибірки.

в) Знаходимо аргумент значення інтегральної функції Лапласа $\gamma = 2\Phi(t) = 0,9545$ по таблиці: $t = 2$. У формулу $\Delta = t\mu$ при $\Delta = 0,03$ підставимо усі відомі величини: $0,03 = 2\sqrt{\frac{0,9 \cdot 0,1}{n}}$, звідси $n = 400$ для повторної вибірки; $0,03 = 2\sqrt{\frac{0,9 \cdot 0,1}{n}\left(1 - \frac{n}{8000}\right)}$, звідси $n = 381$ для безповторної вибірки.

г) Якщо про частку нестандартних телевізорів нічого не відомо, то приймаємо добуток pq рівним максимальному можливому значенню $0,25$.

Для повторної вибірки:

а) $\mu \approx \sqrt{\frac{0,25}{800}} = 0,0177$, а довірча імовірність $P\left(\left|w - p\right| \le 0,02\right) \ge 2\Phi(\frac{0,02}{0,0177}) = 2\Phi(1,13) = 0,7415$;

б) $\Delta = 1,96 \cdot 0,0177 = 0,0347$, а довірчий інтервал: $0,5 - 0,0347 \le p \le 0,5 + 0,0347$ або $0,4653 \le p \le 0,5347$ або $p = 0,5 \pm 0,0347$ або $p = 50\% \pm 3,47\%$;

в) $0,03 = 2\sqrt{\frac{0,25}{n}}$, звідси $n = 1112$.

Для безповторної вибірки:

а) $\mu \approx \sqrt{\frac{0,25}{800}\left(1 - \frac{800}{8000}\right)} = 0,0168$, а надійність $P\left(\left|w - p\right| \le 0,02\right) \ge 2\Phi(\frac{0,02}{0,0168}) = 2\Phi(1,19) = 0,766$;

б) $\Delta = 1,96 \cdot 0,0168 = 0,0329$, а довірчий інтервал: $0,5 - 0,0329 \le p \le 0,5 + 0,0329$ або $0,4671 \le p \le 0,5329$ або $p = 0,5 \pm 0,0329$ або $p = 50\% \pm 3,29\%$;

в) $0,03 = 2\sqrt{\frac{0,25}{n}\left(1 - \frac{n}{8000}\right)}$, звідси $n = 976$.

Розділ 9. Статистичні гіпотези

Часто необхідно знати закон розподілу ознаки у генеральній сукупності. Наприклад, є підстави вважати, що він має вигляд А. Тоді висувають гіпотезу (припущення): **генеральна сукупність розподілена за законом А.** У цій гіпотезі йде мова про вигляд невідомого закону розподілу. Іноді закон розподілу генеральної сукупності відомий, але його параметри (числові характеристики) невідомі. Тоді висувають гіпотезу: **невідомий параметр θ дорівнює θ_o.** Ця гіпотеза вказує припущену величину параметра відомого розподілу. Можливі інші гіпотези: про рівність параметрів двох різних розподілів, про незалежність вибірок тощо.

Означення. **Статистичними** називають гіпотези про вигляд розподілу генеральної сукупності або про параметри відомих розподілів.

Наприклад, статистичними будуть гіпотези: а) генеральна сукупність розподілена за нормальним законом; б) дисперсії двох сукупностей, розподілених за законом Пуассона, рівні між собою.

Приклад нестатистичної гіпотези (оскільки не йде мова ні про вигляд закону розподілу, ні про його параметри): значна частина людей, народжених у другому півріччі, має краще розвинену праву частину мозку, яка здійснює образне мислення.

Разом із припущеною гіпотезою завжди можна розглядати протилежну їй гіпотезу, які доцільно розрізняти.

Означення. **Основною (нульовою)** називають припущену гіпотезу і позначають H_o.

Означення. **Альтернативною (конкурентною)** називають гіпотезу, що суперечить основній і позначають H_1.

Наприклад, якщо $H_o: M(X) = 6$, то $H_1: M(X) \neq 6$.

Гіпотези можуть містити тільки одне припущення (**прості**) або більше одного припущення (**складні**). Наприклад, якщо λ - параметр показникового розподілу, то гіпотеза $H_o: \lambda = 6$ - проста, а гіпотеза $H_o: \lambda > 6$ - складна (містить нескінченну множину гіпотез).

Статистична гіпотеза, яка висунута, може бути правильною або неправильною, тому виникає необхідність її **статистичної перевірки** (перевірка за даними вибірки). При цьому за даними **випадкової вибірки** можна зробити хибний висновок.

Означення. **Якщо за висновком буде відкинута правильна гіпотеза, то кажуть, що це похибка першого роду.**

Означення. **Якщо за висновком буде прийнята хибна гіпотеза, то кажуть, що це похибка другого роду.**

Відмітимо, що наслідки похибок другого роду більш небезпечні, ніж наслідки похибок першого роду.

Означення. Імовірність здійснити похибку першого роду називають **рівнем значущості.**

Рівень значущості найчастіше позначають α і приймають рівним 0,01 або 0,05. Якщо $\alpha = 0{,}05$, то це значить, що в п'яти випадках із 100 ми ризикуємо дістати похибку першого роду (відкинути правильну гіпотезу).

Наприклад, при контролі якості продукції імовірність признати неякісними якісні вироби називають **ризиком виробника**, а імовірність признати якісними неякісні вироби називають **ризиком споживача.**

КРИТЕРІЇ УЗГОДЖЕННЯ ДЛЯ ПЕРЕВІРКИ ГІПОТЕЗ.

Означення. **Статистичним критерієм узгодження перевірки гіпотези (або просто критерієм)** називають випадкову величину **К** (вибіркову функцію), розподіл якої (точний або наближений) відомий і яка застосовується для перевірки основної гіпотези.

Зауваження. Якщо статистична характеристика (вибіркова функція) розподілена нормально, то критерій позначають не буквою **К**, а літерою **Z** (а процес перевірки **Z-тестуванням**). Якщо статистична характеристика розподілена за законом Фішера-Снедекора, то її позначають **F (F–тестування)**. У випадку розподілу статистичної характеристики за законом Стьюдента її позначають **t (t–тестування)**, а у випадку закону "хі-квадрат" - χ^2 (χ^2**-тестування**).

Означення. **Спостереженим значенням критерію узгодження** називають значення відповідного критерію, обчислене за даними вибірки.

Означення. **Критичною областю** називають множину можливих значень критерію, при яких основна гіпотеза відхиляється. Є однобічні та двобічні критичні області.

Означення. **Областю прийняття гіпотези (областю допустимих значень)** називають множину можливих значень критерію, при яких основна гіпотеза приймається.

Для знаходження критичних областей (та областей прийняття гіпотез) задають рівень значущості α, визначають кількості ступенів вільності (це поняття буде розглянуто далі), а потім шукають критичну точку $k_{кр}$ із умови $P(K > k_{kp}) = \alpha$ у випадку правобічної критичної області. Ця точка відокремлює критичну область від області прийняття гіпотези.

<u>Зауваження</u>. Єдиним способом одночасного зменшення імовірностей похибок першого та другого роду є збільшення об'єму вибірки.

КРИТЕРІЙ УЗГОДЖЕННЯ ПІРСОНА (χ^2 -КРИТЕРІЙ).

Критерій Пірсона ефективно використовують для перевірки гіпотези про розподіл генеральної сукупності (теоретичний розподіл). Критерієм перевірки основної гіпотези про вигляд теоретичного розподілу беруть випадкову величину χ^2, що визначається через порівняння емпіричних (вибіркових) та теоретичних частот. Ця ВВ не залежить від виду закону, а залежить тільки від рівня значущості α та кількості ступенів вільності k, яка визначається як різниця між зменшеною на одиницю кількістю варіант x_i (або інтервалів варіант) та кількістю параметрів розподілу. Тобто $k = n - 1 - l$, де n - кількість варіант (або інтервалів варіант), а l - кількість параметрів розподілу.

Критичне значення (критична точка) $\chi^2_{kp} = \chi^2(\alpha; k)$ знаходиться за відповідними таблицями (або за спеціальними функціями Excel).

<u>Правило Пірсона.</u> Щоб при заданому рівні значущості α перевірити основну гіпотезу H_o: генеральна сукупність розподілена за певним законом, потрібно:

1) **припустити наявність певного закону розподілу, знайти його параметри та побудувати (записати) цей закон;**
2) **обчислити за цим законом теоретичні частоти m_i^T для кожної варіанти x_i (або інтервалу варіант);**
3) **обчислити спостережене значення критерію за формулою**

$$\chi^2_{сп} = \sum_{i=1}^{k} \frac{\left(m_i - m_i^T\right)^2}{m_i^T};$$

4) **знайти за таблицями критичну точку $\chi^2_{kp} = \chi^2(\alpha; k)$;**
5) **порівняти χ^2_{kp} та $\chi^2_{сп}$ і зробити висновок:**

якщо $\chi^2_{сп} < \chi^2_{kp}$, то гіпотеза приймається,

якщо $\chi^2_{сп} > \chi^2_{kp}$, то гіпотеза відхиляється.

Для нашого приклада (див.аналіз вибіркових даних на попередній лекції) висунемо основну гіпотезу H_o: **платня усіх робітників фірми (генеральна сукупність) розподілена за нормальним законом з параметрами**:

математичне сподівання $a = M(X) = X_{\tilde{A}} \approx X_B = 325,8$ грн.

стандарт $\sigma = \sigma_{\tilde{A}} \approx \sigma_B = 22,67$ грн.

Щільність розподілу імовірностей (диференціальна функція):

$$\phi(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}} \approx \frac{1}{22,67\sqrt{2\pi}} e^{-\frac{(x-325,8)^2}{1027,72}}.$$

Функція розподілу імовірностей (інтегральна функція):

$F(x) = \frac{1}{2} + \Phi\left(\frac{x - 325,8}{22,67}\right)$, де $\Phi(t)$ - інтегральна функція Лапласа.

Усі подальші розрахунки проводяться за допомогою спеціальних функцій Excel для рівня значущості $\alpha = 0,05$ при кількості ступенів вільності $k = 7 - 1 - 2 = 4$: $\chi^2_{êðèò} = \chi^2(\alpha = 0,05; k = 4) = 9,49$.

ВИСНОВОК: оскільки $\chi^2_{ðîçð} = 9,14 < \chi^2_{êðèò} = 9,49$, то гіпотеза H_o: платня усіх робітників фірми (генеральна сукупність) розподілена за нормальним законом приймається.

Розділ 10. Функціональна, статистична та кореляційна (регресійна) залежності. Проста лінійна регресія

Поняття статистичної та кореляційної залежності

Нагадаємо, що ***функціональна залежність*** характеризується відповідністю кожному значенню однієї змінної (аргумента) цілком певного, єдиного значення іншої змінної (функції).

Означення. Статистичною залежністю між двома змінними називається залежність, при якій кожному можливому значенню однієї змінної відповідає закон розподілу іншої змінної.

Означення. Кореляційною (регресійною) називають залежність, при якій кожному можливому значенню однієї змінної відповідає середнє (умовне середнє) значення іншої змінної (знайдене по закону розподілу або отримане шляхом спостережень). Кореляція – взаємозв'язок, регресія – вплив.

Розглянемо наступний **приклад:**

Залежність між випуском продукції **У** (тон) протягом доби та величиною основних виробничих фондів (ОВФ) **Х** (млн.грн.) для сукупності 50 однотипних підприємств наведена в таблиці :

У / **Х**	27--31	31--35	35--39	39--43	43--47	m_x
40-45	2	1				3
45-50	3	6	4			13
50-55		3	11	7		21
55-60		1	2	6	2	11
60-65				1	1	2
m_y	5	11	17	14	3	50

Необхідно:

А) побудувати точкову діаграму статистичної залежності (кореляційне поле); визначити аргументи (регресори), які впливають на функцію-регресант;

Б) побудувати моделі регресійної залежності. Оцінити щільність кореляційного зв'язку;

В) використати моделі для економічного аналізу та прогнозування.

Спочатку «дискретизуємо» ВВ **Х** та **У**. Для цього кожен інтервал зміни ВВ замінимо на середнє значення. Задано двовимірну статистичну сукупність – **кореляційну таблицю**, графічне зображення якої – **кореляційна хмара** дозволяє зробити припущення про наявність залежності між змінними. Із економічної постановки задачі слідує, що незалежною змінною (регресором) є ОВФ – **Х**, яка впливає на регресант (залежну змінну) - випуск **У**. За даними таблиці побудовано **кореляційну (регресійну)** залежність та її графік – **емпіричну лінію регресії**. Вигляд цієї ламаної теж дозволяє припускати лінійну залежність між змінними.

Проста вибіркова лінійна регресія

Прості лінійні регресійні моделі встановлюють лінійну залежність між двома змінними, наприклад витратами на відпустку та складом родини; витратами на рекламу та обсягом реалізованої продукції; витратами на споживання та валовим національним продуктом (ВНП); зміною обсягу реалізованої продукції залежно від часу тощо.

При цьому одна із змінних вважається залежною (y - ендогенна або результативна змінна, регресант) та розглядається як функція від незалежної змінної (x - екзогенна або факторна змінна, регресор).

У загальному вигляді проста вибіркова регресійна модель запишеться так:

$$y = a_0 + a_1 x + e,$$

де

y - вектор спостережень за залежною змінною; $y = \{y_1, ..., y_n\}$;

x - вектор спостережень за незалежною змінною; $x = \{x_1, ..., x_n\}$;

a_0, a_1 - невідомі параметри регресійної моделі;

e - вектор випадкових величин (помилок); $e = \{e_1, ..., e_n\}$.

Регресійна модель називається лінійною, якщо вона лінійна за своїми параметрами. Її ще можна трактувати як пряму на площині, де a_0 - перетин з віссю ординат, а a_1 - нахил (звичайно, якщо абстрагуватись від випадкової величини e).

Оцінка параметрів лінійної регресії за допомогою методу найменших квадратів (МНК)

Щоб мати явний вигляд залежності, необхідно знайти (оцінити) невідомі параметри a_0, a_1 цієї моделі. Тобто, потрібно за певним критерієм вибрати із множини можливих прямих «найкращу» з точки зору даного критерію. Найпоширенішим є **критерій найменших квадратів**, який полягає у мінімізації суми квадратів відхилень (помилок, залишків)

$$e_i = y_i - y_i^t = y_i - a_0 - a_1 x_i \quad , \quad i = 1,2,...,n.$$

За цим критерієм:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - a_0 - a_1 x_i)^2 = f(a_0, a_1) \to \min.$$

Визначимо значення a_0 , a_1 , які мінімізують суму квадратів відхилень, із необхідних умов екстремуму функції двох змінних (неважко переконатись у виконанні достатніх умов мінімуму для цієї стаціонарної точки). Як відомо, це умови рівності нулю усіх частинних похідних:

$$\begin{cases} f'_{a_0} = \dfrac{\partial\left(\sum e_i^2\right)}{\partial a_0} = -2\sum_{i=1}^{n}\left(y_i - a_0 - a_1 x_i\right) = 0 \\ f'_{a_1} = \dfrac{\partial\left(\sum e_i^2\right)}{\partial a_1} = -2\sum_{i=1}^{n} x_i\left(y_i - a_0 - a_1 x_i\right) = 0 \end{cases}$$

Після нескладних перетворень звідси дістаємо систему лінійних алгебраїчних рівнянь (так звану **нормальну систему**):

$$\begin{cases} n \cdot a_0 + \sum_{i=1}^{n} x_i \cdot a_1 = \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_i \cdot a_0 + \sum_{i=1}^{n} x_i^2 \cdot a_1 = \sum_{i=1}^{n} x_i y_i \end{cases}$$

Розв’язок нормальної системи відносно нахилу a_1 дає

$$a_1 = \frac{\sum x_i y_i - \frac{1}{n}\sum x_i \sum y_i}{\sum x_i^2 - \frac{1}{n}\left(\sum x_i\right)^2}.$$

Поділивши чисельник і знаменник на n, отримуємо

$$a_1 = \frac{\frac{1}{n}\sum x_i y_i - \frac{1}{n^2}\sum x_i \sum y_i}{\frac{1}{n}\sum x_i^2 - \frac{1}{n^2}\left(\sum x_i\right)^2} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\mathbf{var}(x)} = \frac{\mathbf{cov}(x,y)}{\mathbf{var}(x)},$$

де $\bar{z}$ - середні значення, $\mathbf{cov}(x,y)$ - коефіцієнт коваріації, $\mathbf{var}(x)$ - дисперсія.

Враховуючи, що сума відхилень дорівнює нулю ($\sum e_i = 0$), а також знайдене значення a_1, дістаємо значення іншого параметра a_0:

$$a_0 = \bar{y} - a_1 \bar{x}.$$

У нашому прикладі система нормальних рівнянь розв'язана матричним методом.

Властивості простої вибіркової лінійної регресії

1) ***Регресійна пряма проходить через середню точку (це аналогічно тому, що сума помилок дорівнює нулю).***
2) ***Залишки e_i мають нульову коваріацію як зі спотережуваними значеннями x_i, так і з оціненими значеннями $y_i^t = a_0 + a_1 x_i$.***
3) ***Сума квадратів залишків є функцією від кута нахилу a_1.***

Коефіцієнт кореляції

Після знаходження оцінок невідомих параметрів регресійної моделі оцінимо щільність зв'язку між величинами, тобто потрібно відповісти на запитання, наскільки значним є вплив незалежної змінної (фактору, регресора) x на залежну змінну (результат, регресант) y. Найпростішим критерієм, який дає кількісну оцінку зв'язку між двома показниками, є коефіцієнт кореляції:

$$r_{yx} = \frac{\mathbf{cov}(x,y)}{\sqrt{\mathbf{var}(x)\,\mathbf{var}(y)}} = \frac{\mathbf{cov}(x,y)}{\sigma_x \sigma_y},$$

де $\mathbf{cov}(x, y)$ - коефіцієнт коваріації між x та y ; $\mathbf{var}(x), \mathbf{var}(y)$ - дисперсії змінних.

Як видно із виразу, коефіцієнт кореляції, на відміну від коефіцієнта коваріації, є вже не абсолютною, а відносною мірою зв'язку між двома факторами. Тому значення коефіцієнта кореляції розташовані між -1 та +1 ($-1 \le r_{xy} \le 1$). Позитивне значення коефіцієнта кореляції свідчить про прямий зв'язок між факторами, а негативне – про зворотний зв'язок. Коли коефіцієнт кореляції прямує за абсолютною величиною до 1, це свідчить про наявність сильного зв'язку ($r_{xy} \to \pm 1$ - щільність зв'язку велика), коли коефіцієнт кореляції прямує до нуля ($r_{xy} \to 0$), то зв'язок дуже слабкий. У нашому прикладі щільність прямого зв'язку між факторами велика, оскільки коефіцієнт кореляції близький до одиниці.

Декомпозиція дисперсій. Коефіцієнт детермінації

Поряд із коефіцієнтом кореляції використовується ще один критерій, за допомогою якого також вимірюється щільність зв'язку між двома або більше показниками та перевіряється адекватність (відповідність) побудованої регресійної моделі реальній дійсності (фактичним даним). Тобто дається відповідь на запитання, на скільки зміна значень y лінійно залежить від зміни значень x, а не відбувається під впливом різних випадкових факторів, не врахованих у моделі. Таким критерієм є **коефіцієнт детермінації**.

Спочатку розглянемо питання про декомпозицію дисперсій (так зване «правило складання дисперсій»), яке є одним із центральних у статистиці.

Розглянемо на рисунке, як розбиваються на дві частини відхилення фактичних (емпіричних) значень залежної змінної y_i від значень, які знаходяться на регресійній прямій (теоретичних y_i^t або розрахункових y_i^p):

Як видно із рисунка: $(y_i^t - y_i) = (y_i^t - \bar{y}) + (\bar{y} - y_i)$. Звідси дістаємо

$$(y_i - \bar{y}) = (y_i^t - \bar{y}) + (y_i - y_i^t). \qquad (*)$$

В статистиці різницю $(y_i - \bar{y})$ прийнято називати **загальним відхиленням**. Різницю $(y_i^t - \bar{y})$ називають **відхиленням, яке можна пояснити, виходячи**

із регресійної прямої. Різницю $(y_i - y_i^t)$ називають відхиленням, яке не можна пояснити, виходячи з регресійної прямої, або **непояснюваним відхиленням**. Піднесемо обидві частини (*) до квадрату і підсумуємо по i. Враховуючи, що сума похибок дорівнює нулю, дістанемо:

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i^t - \bar{y})^2 + \sum_{i=1}^{n}(y_i - y_i^t)^2, \qquad (**)$$

де $\sum_{i=1}^{n}(y_i - \bar{y})^2$ - загальна сума квадратів, яку прийнято позначати **SST (sum square total)**; $\sum_{i=1}^{n}(y_i^t - \bar{y})^2$ - сума квадратів, що пояснює регресію та позначається **SSR (sum square regression)**; $\sum_{i=1}^{n}(y_i - y_i^t)^2$ - сума квадратів помилок, яка позначається **SSE (sum square error)**. Таким чином, (**) у скороченому вигляді може бути записана як

SST=SSR+SSE.

Поділивши обидві частини (*) на n, отримаємо так зване «правило складання дисперсій»:

$$\sigma_{заг}^2 = \sigma_{регр}^2 + \sigma_{пом}^2, \qquad (***)$$

де $\frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{y})^2$ - загальна дисперсія, яка позначена $\sigma_{заг}^2$;

$\frac{1}{n}\sum_{i=1}^{n}(y_i^t - \bar{y})^2$ - дисперсія, що пояснює регресію, позначається $\sigma_{регр}^2$;

$\frac{1}{n}\sum_{i=1}^{n}(y_i - y_i^t)^2$ - дисперсія помилок, яка позначена $\sigma_{пом}^2$.

Таким чином, ми розклали загальну дисперсію на дві частини: дисперсію, що пояснює регресію, та дисперсію помилок (або дисперсію випадкової величини).

Поділимо обидві частини (***) на $\sigma_{заг}^2$ і отримаємо:

$$1 = \frac{\sigma^2_{регр}}{\sigma^2_{заг}} + \frac{\sigma^2_{пом}}{\sigma^2_{заг}}.$$

Як видно, перше відношення у правій частині є пропорцією дисперсії, що пояснює регресію, у загальній дисперсії. Друге відношення є пропорцією дисперсії помилок у загальній дисперсії, тобто є частиною дисперсії, яку не можна пояснити через регресійний зв'язок.

Частина дисперсії, що пояснює регресію, називається коефіцієнтом детермінації і позначається R^2. Коефіцієнт детермінації використовується як критерій адекватності моделі, оскільки є мірою пояснювальної сили незалежної змінної x. Коефіцієнт детермінації можна записати в одному із двох еквівалентних виразів:

$$R^2 = \frac{\sigma^2_{регр}}{\sigma^2_{заг}} \text{ або } R^2 = \frac{SSR}{SST}.$$

Очевидно, що $0 \le R^2 \le 1$.

Враховуючи, що коефіцієнт кореляції $r = a_1 \frac{\sigma_x}{\sigma_y}$, неважко встановити наступний зв'язок між коефіцієнтами детермінації та кореляції (для лінійної регресії):

$$R^2 = r^2.$$

Поняття про ступені вільності.

Повернемося до тотожності, яка зв'язує суми квадратів:

SST=SSR+SSE.

Кожна сума квадратів пов'язана з числом, яке називають її «**ступенем вільності**». Це число показує, скільки незалежних елементів інформації, що утворилися із спостережуваних елементів $y_1, y_2, \ldots, y_n$, потрібно для розрахунку даної суми квадратів.

У статистиці кількістю ступенів вільності певної величини називають різницю між кількістю різних дослідів та кількістю

параметрів, встановлених у результаті цих дослідів, незалежно один від одного.

Розглянемо, скільки ступенів вільності має кожна сума квадратів.

Загальна сума квадратів SST утворюється із використанням $(n-1)$ **незалежних чисел, тому що із чисел** $\{(y_1-\bar{y}),(y_2-\bar{y}),...,(y_n-\bar{y})\}$ незалежні тільки $(n-1)$ враховуючи властивість $\sum_{i=1}^{n}(y_i-\bar{y})=0$.

Суму квадратів, що пояснює регресію (SSR), отримують, використовуючи тільки одну незалежну одиницю інформації, яка утворюється із $y_1, y_2, ..., y_n$ **, а саме** a_1 **(для випадку багатофакторної регресії матимемо іншу ситуацію).** Звідси **SSR** має один ступінь вільності. Звернемо увагу на те, що кількість ступенів вільності співпадає із кількістю незалежних змінних, що входять до регресійної моделі.

Сума квадратів помилок (SSE) має $(n-2)$ **ступені вільності.** Ця сума базується на кількості ступенів вільності, яка дорівнює різниці між кількістю спостережень та кількістю параметрів, що оцінюються.

Ступені вільності прийнято позначати DF , або Df , або df .

У разі простої лінійної регресії ступені вільності можна розкласти як суми квадратів: $n-1=1+(n-2)$.

Перевірка простої регресійної моделі на адекватність.

Поняття F-критерію Фішера.

Ми показали, що адекватність простої лінійної регресії можна перевірити за допомогою коефіцієнта детермінації. Якщо його значення близьке до одиниці, то модель адекватна. Якщо його значення близьке до нуля, то модель неадекватна. Проблема оцінки адекватності виникає, коли коефіцієнт детермінації набуває «проміжних значень», напр. 0,3; 0,5; 0,7 тощо. У таких випадках важко зробити однозначний висновок щодо адекватності моделі, тому потрібен відповідний критерій. Найпоширенішим із таких критеріїв є критерій Фішера. До правої частини простої лінійної регресійної моделі $y_i = a_0 + a_1 x_i + e_i$ входить випадкова величина e_i, тому величини y_i будуть також випадковими, як і будь-які функції від них.

В теорії імовірностей розглядається величина

$$F = \frac{MSR}{MSE},$$

де **mean square regression** $MSR = SSR/1$ - середня сума квадратів, що пояснює регресію (тобто сума квадратів, поділена на відповідний ступень вільності), **mean square error** $MSE = SSE/(n-2)$ - середня сума помилок. Як відомо ця величина має функцію розподілу Фішера $F\{1; n-2\}$ із 1 та $(n-2)$ ступенями вільності, за умови, що нахил лінійної моделі дорівнює нулю. На цьому базується F -критерій Фішера, процес застосування якого можна поділити на наступні етапи:

1) Розраховуємо величину F -відношення

$$F_{\{1;n-2\}} = \frac{MSR}{MSE}.$$

2) Задаємо рівень значимості (значущості) α (або $\alpha \cdot 100\%$). Наприклад, якщо ми вважаємо, що можлива для нас помилка становить $\alpha = 0{,}05$ (або 5%), це означає, що ми можемо помилитись не більше, ніж у 5% випадків, а в 95% випадків наші висновки будуть правильними.

3) За статистичними таблицями F -розподілу Фішера з $\{1; n-2\}$ ступенями вільності і рівнем значимості α (або $1-\alpha$) обчислюємо критичне значення $F_{кр}$.

4) Якщо розраховне значення $F > F_{кр}$, то ми відкидаємо нульову (базову) гіпотезу, що нахил нульовий з ризиком помилитися не більше ніж у 5% випадків. Іншими словами у цьому випадку побудована нами модель адекватна реальній дійсності. У супротивному випадку ($F \le F_{кр}$) модель неадекватна.

Розглянемо застосування критерію Фішера у нашому прикладі. Висновок: із 95% надійністю побудована модель адекватна вибірковим даним.

Прогнозування.

Після побудови моделі (теоретичної регресійної залежності) та перевірки її адекватності можна виконувати прогнозування. При цьому отримуємо точкові та інтервальні прогнози. Точковий прогноз дає оцінку значення залежної змінної, наприклад, для значення x_{n+1} за побудованою вибірковою моделлю: $y^t_{n+1} = a_0 + a_1 x_{n+1}$.

При інтервальному оцінюванні застосовується t-розподіл Стьюдента з $(n-2)$ ступенями вільності при заданому рівні значущості α:

$$y = y^t_{n+1} \pm t_{\alpha/2} \cdot \sigma_\varepsilon \cdot \sqrt{\left(1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)},$$

де σ_ε - середня квадратична помилка, $\bar{x}$ - середнє значення фактору (регресора). Для нашого прикладу знайдено 95% довірчий інтервал прогнозованого випуску при збільшенні ОВФ до 70 млн.грн.

Економічна інтерпретація: коефіцієнт при змінній X в моделі означає, що при збільшенні ОВФ на 1 млн.грн випуск зросте приблизно на 0,646 тон.

Класична модель лінійної регресії: основні припущення, що лежать в основі методу найменших квадратів

Узагальнена лінійна регресійна модель

$$y = \beta_0 + \beta_1 x + \varepsilon,$$

де β_0, β_1 - правильні параметри усієї генеральної сукупності, ε - неспостережувана випадкова величина.

Мета регресійного аналізу полягає не тільки у визначенні невідомих параметрів вибіркової лінійної моделі a_0, a_1, а, насамперед, у висновках, які ми можемо зробити щодо дійсних значень параметрів узагальненої моделі β_0, β_1. Для того, щоб відповісти на запитання, наскільки наближаються знайдені оцінки a_0, a_1 до відповідних значень параметрів узагальненої моделі, або, що те ж саме, наскільки наближається теоретичне значення y_i^t до дійсного значення свого математичного сподівання $E\left(y / x_i\right)$, ми повинні не тільки точно визначити функціональну форму моделі, а й зробити певні припущення щодо випадкової величини (ВВ) ε та зв'язку між випадковою величиною та незалежною змінною x_i.

Припущення 1. Математичне сподівання ВВ ε дорівнює нулю:

$$E\left(\varepsilon_i / x_i\right) = 0.$$

Припущення 1 реально стверджує, що фактори, які не враховано в моделі і тому віднесено до ε_i, не впливають систематично на математичне сподівання y, тобто додатні значення ε_i нейтралізують від'ємні ε_i, тому їхній усереднений чи очікуваний вплив на y дорівнює нулю.

Зазначимо, що припущення $E\left(\varepsilon_i / x_i\right) = 0$ еквівалентне умові $E\left(y_i / x_i\right) = \beta_0 + \beta_1 x_i$.

Припущення 2. Відсутність автокореляції між ВВ ε_i. Це припущення означає, що ВВ ε_i незалежні між собою, тобто коефіцієнт коваріації між ними дорівнює нулю:

$$cov(\varepsilon_i, \varepsilon_j) = E\left(\left[\varepsilon_i - E(\varepsilon_i)\right]\left[\varepsilon_j - E(\varepsilon_j)\right]\right) = E(\varepsilon_i \varepsilon_j) = 0.$$

Припущення 2 дає змогу розглянути найпростіший випадок, коли вивчається систематичний вплив (якщо він є) x_t на y_t без урахування впливу інших факторів, виражених ВВ ε. У супротивному випадку y_t залежатиме не тільки від x_t, а й від ε_{t-1}. Тому потрібно тестувати наявність зв'язку між ВВ ε.

Припущення 3. Гомоскедастичність, або однакова дисперсія ВВ ε_i незалежно від номера спостереження:

$$var(\varepsilon_i / x_i) = E\left(\left[\varepsilon_i - E(\varepsilon_i)\right]^2\right) = E(\varepsilon_i^2) = \sigma^2.$$

Це припущення свідчить, що умовна дисперсія розподілу y є також сталою величиною.

Припущення 4. Незалежність між значеннями ε_i і значеннями змінної x_i, або нульова коваріація між ними:

$$cov(\varepsilon_i, x_i) = E\left(\left[\varepsilon_i - E(\varepsilon_i)\right]\left[x_i - E(x_i)\right]\right) = E(\varepsilon_i x_i) = 0.$$

Припущення 4 виконується автоматично, якщо змінна x є невипадковою, або нестохастичною (як у нашому прикладі ряду динаміки). Дане припущення суттєве у випадку, коли значення x випадкові.

Припущення 5. Регресійну модель визначено (специфіковано) правильно (відсутність похибки). Покажемо, що може бути у випадку порушення цього припущення на прикладі так званої кривої Філіпса.

ПРАКТИЧНЕ ЗАНЯТТЯ №1

1. **Елементи комбінаторики.**
2. **Класифікація подій.**
3. **Класичне та статистичне означення імовірності.**
4. **Алгебра подій.**
5. **Теорема добутку.**

ЗАДАЧІ ДЛЯ РОЗВ'ЯЗУВАННЯ В АУДИТОРІЇ.

Приклад 1.1. Дана множина $A = \{a; b; c\}$. Випишіть усі можливі перестановки, розміщення та комбінації. Підрахуйте їх кількість та порівняйте із розрахунками за формулами.

Приклад 1.2. Опишіть простір елементарних подій для наступних випробувань: а) подвійне підкидання монети; б) постріли по мішені до першого влучення.

Приклад 1.3. Стрілець стріляє двічі по мішені. Опишіть простір елементарних наслідків. Виберіть із цього простору наступні події: а) стрілець влучив у мішень принаймні один раз; б) стрілець влучив тільки один раз; в) стрілець не влучив у мішень.

Приклад 1.4. На складі є деталі трьох сортів. Навмання вибирається деталь. Розглядаються події: А - деталь першого сорту; В - деталь другого сорту; С - деталь третього сорту. Опишіть події $A+B$, $\overline{A+B}$, AC, $AB+C$.

Приклад 1.5. З колоди у 36 карт навмання беруть одну карту. Знайти ймовірності появи наступних подій: А – вийнята карта бубнової масті, В - вийнята карта туз, С-вийнята карта бубнова сімка.

Приклад 1.6. На складі зберігають 500 акумуляторів. Відомо, що після року зберігання 4% із них будуть непридатними. Знайти імовірність того, що навмання взятий після року зберігання акумулятор буде справним, якщо відомо, що після 6 місяців зберігання було вилучено 5 несправних акумуляторів.

Приклад 1.7. Відділ контролю виявив 5 бракованих виробів із випадково відібраних 100 однакових виробів. Знайти частість появи бракованих виробів.

Приклад 1.8. В урні знаходяться 4 білі та 6 чорних куль. Яка ймовірність того, що навмання витягнуті 2 кулі будуть: а) чорні? б) білі? (Розгляньте два випадки: діставання куль виконується за схемами "повернених" та "неповернених" куль).

Приклад 1.9. Студент підготував 20 із 30 теоретичних питань, які входять до екзаменаційного білету. Знайти імовірність того, що навмання взятий білет на екзамені (із двома теоретичними питаннями) буде містити принаймні одне питання, підготовлене студентом.

Приклад 1.10. У студентскій групі 25 чоловік. Серед них 20 – старші 19 років і 8 – старші 22 років. Знайти імовірність того, що навмання вибраний із групи студент старший 19 років, але не старший 22 років.

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ.

Приклад 1.11. Дана множина $A=\{+;\ -;\ \times;\ \div\}$. Випишіть усі можливі перестановки, розміщення та комбінації. Підрахуйте їх кількість та порівняйте із розрахунками за формулами.

Приклад 1.12. Кидається гральний кубик. Опишіть простір елементарних подій. Опишіть подію А – випадіння числа, що ділиться на 3.

Приклад 1.13. Гральний кубик кидають двічі. Опишіть простір елементарних подій. Опишіть події: А – сума очок, яка випаде на двох кубиках, дорівнює 8; В – випаде принаймні одна 6.

Приклад 1.14. В урні є 10 куль : 3 білі та 7 чорних . Яка ймовірність того, що навмання витягнуті 2 кулі будуть: а) чорні? б) білі?

Приклад 1.15. Вкладники банку за сумами вкладів та віком мають такий процентний розподіл:

Вік	Суми вкладу		
	<$1000	$1000-5000	>$ 5000
< 30 років	5%	15%	8%
30 – 50 років	8%	25%	20%
> 50 років	7%	10%	2%

Нехай А та В – такі події:

А={у навмання вибраного клієнта вклад більший $5000}

В={вік навмання вибраного клієнта не менший 30 років}.

Визначити: P(A),P(B), $P(A\cup B), P(A\cap B)$.

ПРАКТИЧНЕ ЗАНЯТТЯ №2

1. Теореми добутку (продовження) та суми.

2. Повна імовірність.

3. Формула Байєса.

ЗАДАЧІ ДЛЯ РОЗВ'ЯЗУВАННЯ В АУДИТОРІЇ.

Приклад 2.1. Два мисливці роблять по одному пострілу у ціль. Імовірність влучання для першого мисливця становить 0,7 , а для другого – 0,8. Знайти імовірність того, що: а) обидва влучать у ціль; б) жоден не влучить; в) хоча б один влучить; г) тільки один влучить.

Приклад 2.2. Знайдіть імовірність $P(A)$, якщо $P(AB) = 0{,}74 \quad ; \quad P(A\overline{B}) = 0{,}16$.

Приклад 2.3. Імовірність хоча б одного влучення в ціль при трьох пострілах дорівнює 0,992. Знайдіть імовірність влучення при одному окремому пострілі.

Приклад 2.4. Серед 30 екзаменаційних білетів є 5 "щасливих". Студенти по черзі витягують білети. У кого більша імовірність витягти "щасливий" білет: у того, хто підійшов першим, чи у того, хто підійшов другим?

Приклад 2.5. На двох автоматичних лініях виготовляють однакові деталі, причому 60% на першій та 40% на другій. Імовірність виготовлення стандартної деталі на першій лінії дорівнює 0,8 , а на другій – 0,9. Виготовлені деталі надходять до складу. Знайти імовірність того, що навмання взята зі складу деталь не відповідає стандарту.

Приклад 2.6. У першому ящику вдвічі більше деталей, ніж у другому. Перший містить 20% браку, а другий – 10%. При транспортуванні загубили якісну деталь. Знайти імовірність того, що загублено деталь із : а) першого ящика; б) другого ящика.

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ.

Приклад 2.7. Кидають три гральних кубика. Знайти ймовірність того, що число 6 випаде : а) принаймні на одному із кубиків; б) тільки на одному із кубиків.

Приклад 2.8. Знайдіть імовірність $P(B)$, якщо $P(AB) = 0{,}85 \quad ; \quad P(\overline{A}B) = 0{,}12$..

Приклад 2.9. У першій урні 3 білі та 7 чорних куль, а у другій – 4 білі та 6 чорних куль. Із першої урни до другої переклали дві кулі. Знайти імовірність того, що навмання витягнута із другої урни куля – біла.

Приклад 2.10. До лікарні надходить 50% хворих на грип, 30% хворих на ангіну та 20% хворих на пневмонію. Імовірність повного одужання від грипу дорівнює 0,7 , від ангіни та пневмонії – 0,8 та 0,9 відповідно. Виписано хворого, який повністю одужав. Знайти імовірність того, що він був хворим на: а) ангіну; б) пневмонію; в) грип.

Приклад 2.11. Студент може з однаковою імовірністю зайти до будь-якої із трьох бібліотек. Імовірність отримання потрібної студентові книги становить

для першої бібліотеки 0,9 , для другої – 0,7 , для третьої – 0,6. Студент отримав потрібну книгу. Знайти імовірність того, що книгу видала третя бібліотека.

Приклад 2.12. У першій урні 3 білі та 7 чорних куль, а у другій – 4 білі та 6 чорних куль. Із двох урн дістали по одній кулі. Потім із двох витягнутих куль навмання вибрали одну. Знайти імовірність того, що остання вибрана куля – біла.

ПРАКТИЧНЕ ЗАНЯТТЯ №3

1. Дискретні випадкові величини (ДВВ), їх закони розподілу.

2. Операції над ДВВ.

3. Числові характеристики ДВВ та їх властивості.

4. Метод моментів для обчислення МС та дисперсії.

ЗАДАЧІ ДЛЯ РОЗВ'ЯЗУВАННЯ В АУДИТОРІЇ.

Приклад 3.1. У відділі електротоварів є 5 приладів, серед яких 2 бракованих. Продавець перевіряє їх при продажу, доки не знайде справний прилад. Скласти закон розподілу ВВ – кількості перевірених приладів. Побудувати полігон розподілу імовірностей. Обчислити числові характеристики.

Приклад 3.2. Дослідженнями, проведеними на виробничому об'єднанні за деякий період, встановлені закони розподілу ВВ – об'ємів випуску (в тис.шт.) однорідної продукції для кожного із двох груп підприємств:

а) для кожного підприємства першої групи:

X (тис.шт)	1	2	3
P	0,2	0,5	0,3

б) для кожного підприємства другої групи:

У (тис.шт)	2	3	4
P	0,6	0,3	0,1

Знайти закон розподілу ВВ – об'ємів випуску (в тис.шт.) продукції для всього виробничого об'єднання, якщо підприємств першої групи – 3, а другої групи – 2. Обчислити числові характеристики усіх ВВ (безпосередньо та за властивостями). Побудувати полігони розподілів.

Приклад 3.3. Знайти числові характеристики наступної ВВ:

X	23-32	32-41	41-50	50-59
P	0,1	0,2	0,6	?

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ.

Приклад 3.4. Контролер перевіряє 10 деталей, серед яких є 3 бракованих, до тих пір, поки не знайде якісну. Скласти закон розподілу ВВ – кількості перевірених деталей. Побудувати полігон розподілу імовірностей. Обчислити числові характеристики.

Приклад 3.5. Маркетинговими дослідженнями встановлені закони розподілу ВВ – об'ємів продажу (в тис.шт.) деякого товару та ціни (в грн.) на цей товар протягом певного періоду часу (для спрощення вважати їх незалежними):

об'єми продажу – **X** (тис.шт)	5	7	9
P	0,2	0,5	0,3

ціна одиниці товару – **У** (грн.)	2	3	4
P	0,6	0,3	0,1

Знайти закон розподілу ВВ – виручки (в тис.грн.) від продажу товару протягом досліджуваного періоду. Обчислити числові характеристики усіх ВВ (безпосередньо та за властивостями). Побудувати полігони розподілів.

Приклад 3.6. Використовуючи метод моментів, знайти числові характеристики наступної ВВ:

X	20-25	25-30	30-35	35-40
P	?	0,3	0,4	0,1

ПРАКТИЧНЕ ЗАНЯТТЯ №4

6. **Незалежні повторні випробування (НПВ). Формула Бернуллі.**
7. **Біноміальний закон розподілу (закон Бернуллі).**
8. **Локальна формула Лапласа, формула Пуассона.**
9. **Закон Пуассона (закон рідкісних подій).**

ЗАДАЧІ ДЛЯ РОЗВ'ЯЗУВАННЯ В АУДИТОРІЇ.

Приклад 4.1. В середньому 30% пакетів акцій продаються на аукціоні за початково заявленою ціною. На торги виставлено 20 пакетів. Знайти імовірність того, що за початково заявленною ціною: а) не буде продано жодного пакету; б) буде продано хоча б один пакет; в) скласти закон розподілу частоти проданих пакетів та побудувати полігон; г) знайти найімовірніше число проданих пакетів та його імовірність.

Приклад 4.2. За статистичними даними податкових інспекцій кожне третє із 1000 малих підприємств регіону має порушення фінансової дисципліни. Знайдіть: а) числові характеристики ВВ – частоти малих підприємств, які працюють без порушень; б) найімовірнішу кількість підприємств, які мають

порушення та її імовірність; в) скільки підприємств потрібно перевірити податковим інспекціям, щоб найімовірніша кількість порушників дорівнювала 100?

Приклад 4.3. До банку надійшло 5000 грошових купюр. Імовірність появи фальшивої купюри становить 0,0001. Знайдіть імовірність того, що в результаті перевірки банком буде знайдено 4 фальшивих купюри.

Приклад 4.4. Диспетчерський пункт в середньому протягом хвилини приймає 3 замовлення на таксі. Знайдіть імовірність того, що протягом двох хвилин надійде 4 замовлення.

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ.

Приклад 4.5. В середньому 80% студентів вчасно отримують залік з деякої дисципліни. Знайти імовірність того, що серед випадково вибраних 5 студентів: а) жоден вчасно не отримає залік; б) вчасно отримає залік хоча б один студент; в) скласти закон розподілу кількості студентів, які вчасно отримають залік та побудувати полігон; г) знайти найімовірніше число студентів, які вчасно отримають залік та його імовірність.

Приклад 4.6. За статистичними даними 40% студентів одного із факультетів ОНЕУ палять. На факультеті навчається 1500 студентів. Знайдіть: а) числові характеристики ВВ – числа студентів факультету, які не палять; б) найімовірнішу кількість студентів, які палять та її імовірність; в) скільки студентів факультету потрібно опросити декану, щоб найімовірніша кількість тих, хто палить дорівнювала 50?

Приклад 4.7. Імовірність того, що акція оформлена без порушень дорівнює 0,998. Знайдіть імовірність того, що серед 2000 акцій, які продаються на аукціоні, будуть 5 акцій, оформлених із порушеннями.

Приклад 4.8. На популярну телепередачу в середньому протягом хвилини надходить 10 СМС-повідомлень. Знайдіть імовірність того, що за три години, які триває передача, надійде 5 СМС-повідомлень .

ПРАКТИЧНЕ ЗАНЯТТЯ №5

1. Функція розподілу ймовірностей (інтегральна функція) та її властивості.

2. Щільність розподілу ймовірностей (диференціальна функція) та її властивості.

3. Числові характеристики НВВ.

4. Деякі закони розподілу НВВ (рівномірний, показниковий, нормальний).

ЗАДАЧІ ДЛЯ РОЗВ'ЯЗУВАННЯ В АУДИТОРІЇ.

Приклад 5.1. Скласти закон розподілу ВВ X - числа появи герба при двох кидках монети. Знайти функцію розподілу імовірностей для цієї ВВ та побудувати її графік.

Приклад 5.2. ВВ X рівномірно розподілена на проміжку $[2;7]$. Знайти: а) функції розподілу, побудувати їх графіки; б) числові характеристики; в) імовірності $P(X=3)$, $P(X<4)$, $P(1\le X<5)$.

Приклад 5.3. ВВ X задана функцією розподілу імовірностей

$$F(x)=\begin{cases} 0, & x<1 \\ \dfrac{x-1}{4}, & 1\le x\le 5 \\ 1, & x>5 \end{cases}$$

Знайти: а) щільність розподілу імовірностей, побудувати графіки функцій розподілу; б) числові характеристики; в) імовірності $P(X=4)$, $P(X<3)$, $P(3\le X\le 15)$, показати на графіках числові характеристики та знайдені імовірності.

Приклад 5.4. ВВ X задана щільністю розподілу імовірностей

$$f(x)=\begin{cases} 0, & x<3 \\ const, & 3\le x\le 9 \\ 0, & x>9 \end{cases}$$

Знайти: а) функцію розподілу імовірностей, побудувати графіки функцій розподілу; б) числові характеристики; в) імовірності $P(X=5)$, $P(X>5)$, $P(5\le X<10)$, показати на графіках числові характеристики та знайдені імовірності.

Приклад 5.5. ВВ X розподілена за показниковим законом з параметром $\lambda=2$. Знайти: а) функції розподілу, побудувати їх графіки; б) числові характеристики; в) імовірності $P(X=3)$, $P(X<4)$, $P(1\le X<5)$.

Приклад 5.6. ВВ X задана функцією розподілу імовірностей

$$F(x)=\begin{cases} 0, & x<0 \\ 1-e^{-4x}, & x\ge 0 \end{cases}$$

Знайти: а) щільність розподілу імовірностей, побудувати графіки функцій розподілу; б) числові характеристики; в) імовірності $P(X=4)$, $P(X<3)$, $P(3\le X<15)$, показати на графіках числові характеристики та знайдені імовірності.

Приклад 5.7. ВВ X задана щільністю розподілу імовірностей

$$f(x)=\begin{cases} 0 & , \quad x<0 \\ 0.2e^{-0.2x} & , \quad x\ge 0 \end{cases}$$

Знайти: а) функцію розподілу імовірностей, побудувати графіки функцій розподілу; б) числові характеристики; в) імовірності $P(X=5)$, $P(X>5)$, $P(5\le X<10)$, показати на графіках числові характеристики та знайдені імовірності..

Приклад 5.8. ВВ X розподілена за нормальним законом з параметрами $a=-3$, $\sigma=2$. Знайти: а) функції розподілу, побудувати їх графіки; б) числові характеристики; в) імовірності $P(X=3)$, $P(X<4)$, $P(1\le X<5)$.

Приклад 5.9. ВВ X задана функцією розподілу імовірностей

$F(x)=\frac{1}{2}+\Phi\left(\frac{x+5}{3}\right)$, де $\Phi(t)$ - інтегральна функція Лапласа.

Знайти: а) щільність розподілу імовірностей, побудувати графіки функцій розподілу; б) числові характеристики; в) імовірності $P(X=4)$, $P(X<3)$, $P(3\le X<15)$, показати на графіках числові характеристики та знайдені імовірності.

Приклад 5.10. ВВ X задана щільністю розподілу імовірностей

$$f(x)=\frac{1}{5\sqrt{2\pi}}e^{-\frac{(x-2)^2}{50}}$$

Знайти: а) функцію розподілу імовірностей, побудувати графіки функцій розподілу; б) числові характеристики; в) імовірності $P(X>5)$, $P(1\le X\le 3)$, $P\left(\left|X-M(X)\right|\le 1\right)$, показати на графіках числові характеристики та знайдені імовірності..

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ’ЯЗУВАННЯ.

Приклад 5.11. Скласти закон розподілу ВВ X - числа появи одиниці при двох кидках грального кубика. Знайти функцію розподілу імовірностей для цієї ВВ та побудувати її графік.

Приклад 5.12. ВВ X рівномірно розподілена на проміжку $[-2;7]$. Знайти: а) функції розподілу, побудувати їх графіки; б) числові характеристики; в) імовірності $P(X=-1)$, $P(X<1)$, $P(-1\le X<5)$.

Приклад 5.13. ВВ X задана функцією розподілу імовірностей

$$F(x)=\begin{cases} 0, & x<-1 \\ \dfrac{x+1}{4}, & -1\le x\le 3 \\ 1, & x>3 \end{cases}$$

Знайти: а) щільність розподілу імовірностей, побудувати графіки функцій розподілу; б) числові характеристики; в) імовірності $P(X=4)$, $P(X<3)$, $P(3\le X\le 15)$, показати на графіках числові характеристики та знайдені імовірності.

Приклад 5.14. ВВ X задана щільністю розподілу імовірностей

$$f(x)=\begin{cases} 0, & x<5 \\ const, & 5\le x\le 8 \\ 0, & x>8 \end{cases}$$

Знайти: а) функцію розподілу імовірностей, побудувати графіки функцій розподілу; б) числові характеристики; в) імовірності $P(X=6)$, $P(X>6)$, $P(7\le X<10)$, показати на графіках числові характеристики та знайдені імовірності..

Приклад 5.15. ВВ X розподілена за показниковим законом з параметром $\lambda=5$. Знайти: а) функції розподілу, побудувати їх графіки; б) числові характеристики; в) імовірності $P(X=3)$, $P(X<4)$, $P(1\le X<5)$.

Приклад 5.16. ВВ X задана функцією розподілу імовірностей

$$F(x)=\begin{cases} 0, & x<0 \\ 1-e^{-3x}, & x\ge 0 \end{cases}$$

Знайти: а) щільність розподілу імовірностей, побудувати графіки функцій розподілу; б) числові характеристики; в) імовірності $P(X=4)$, $P(X<3)$, $P(3\le X<15)$, показати на графіках числові характеристики та знайдені імовірності.

Приклад 1.17. ВВ X задана щільністю розподілу імовірностей

$$f(x)=\begin{cases}0, & x<0\\ e^{-x}, & x\ge 0\end{cases}$$

Знайти: а) функцію розподілу імовірностей, побудувати графіки функцій розподілу; б) числові характеристики; в) імовірності $P(X=5)$, $P(X>5)$, $P(5\le X<10)$, показати на графіках числові характеристики та знайдені імовірності..

Приклад 5.18. ВВ X розподілена за нормальним законом з параметрами $a=7$, $\sigma=4$. Знайти: а) функції розподілу, побудувати їх графіки; б) числові характеристики; в) імовірності $P(X=3)$, $P(X<4)$, $P(-1\le X<15)$.

Приклад 5.19. ВВ X задана функцією розподілу імовірностей

$F(x)=\frac{1}{2}+\Phi\left(\frac{x-5}{3}\right)$, де $\Phi(t)$ - інтегральна функція Лапласа.

Знайти: а) щільність розподілу імовірностей, побудувати графіки функцій розподілу; б) числові характеристики; в) імовірності $P(X=4)$, $P(X<3)$, $P(3\le X<15)$, показати на графіках числові характеристики та знайдені імовірності.

Приклад 5.20. ВВ X задана щільністю розподілу імовірностей

$$f(x)=\frac{1}{3\sqrt{2\pi}}e^{-\frac{(x+2)^2}{18}}$$

Знайти: а) функцію розподілу імовірностей, побудувати графіки функцій розподілу; б) числові характеристики; в) імовірності $P(X>5)$, $P(-1\le X\le 3)$, $P\left(\left|X-M(X)\right|\le 1\right)$, показати на графіках числові характеристики та знайдені імовірності..

ПРАКТИЧНЕ ЗАНЯТТЯ №6

1. Закон великих чисел. Нерівності Маркова та Чебишова. Частинні випадки нерівності Чебишова.

2. Збіжність за імовірністю. Теорема Бернуллі. Теорема Чебишова.

3. Центральна гранична теорема.

4. Інтегральна теорема Муавра-Лапласа та її частинні випадки.

ЗАДАЧІ ДЛЯ РОЗВ'ЯЗУВАННЯ В АУДИТОРІЇ

Приклад 6.1. Середня кількість викликів, які потрапляють на міні АТС протягом години, дорівнює 300. Оцініть імовірність того, що протягом деякої години кількість викликів буде: а) не менше 400; б) менше 500.

Приклад 6.2. Практика показує, що 60% студентів II курсу успішно складають усі іспити в період сесії. За допомогою нерівності Чебишова оцініть імовірність того, що серед 1200 студентів II курсу частка тих, хто вчасно здав усі іспити, знаходиться в межах від 0,56 до 0,64.

Приклад 6.3. Ймовірність випуску стандартної деталі дорівнює 0,96. Оцініть за допомогою нерівності Чебишова ймовірність того, що кількість бракованих деталей серед 2000 виготовлених знаходиться у межах від 60 до 100 включно. Уточніть імовірність тієї ж події за допомогою інтегральної теореми Муавра-Лапласа. Порівняйте отримані результати.

Приклад 6.4. На лекції з «Теорії ймовірностей» спізнюються 3 із 75 студентів. Що можна стверджувати про кількість студентів, які спізняться на чергову лекцію, з імовірністю не менше 0,8? Розрахунки провести із застосуванням нерівностей та теореми Муавра-Лапласа.

Приклад 6.5. Середнє квадратичне відхилення кожної із 1000 незалежних однаково розподілених ВВ дорівнює 2. Яке найбільше відхилення середньої арифметичної цих ВВ від свого математичного сподівання за абсолютною величиною можна очікувати із імовірністю, не меншою, ніж 0,95. Порівняти результат із розрахунками за наслідком із центральної граничної теореми.

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Приклад 6.6. Середні витрати електроенергії малим підприємством складає 1000 кВт у день, а середнє квадратичне відхилення цих витрат не перевищує 200 кВт. Оцініть імовірність того, що витрати електроенергії на підприємстві протягом деякого дня не перевищать 2000 кВт, використовуючи : а) нерівність Маркова; б) нерівність Чебишова. Порівняйте отримані оцінки.

Приклад 6.7. Протягом одних біржових торгів курс акцій компанії в середньому змінюється на 0,3%. Оцініть імовірність того, що на наступних торгах курс зміниться більше, ніж на 3%.

Приклад 6.8. У середньому 10% працездатного населення деякого регіону – безробітні. Оцінить за допомогою нерівності Чебишова ймовірність того, що

рівень безробіття (частка безробітних) серед 10000 працездатних жителів міста буде в межах від 9 до 11% (включно).

Приклад 6.9. Досвід роботи страхової компанії показує, що страховий випадок припадає на кожну десяту угоду. Скільки угод необхідно укласти, щоб із імовірністю не менше 0,9 можна було стверджувати, що частка страхових випадків відхилиться від 0,1 не більш ніж на 0,01 (за абсолютною величиною). Порівняйте результат із розрахунками за допомогою наслідку з інтегральної теореми Муавра-Лапласа.

Приклад 6.10. Дисперсія кожної із 900 незалежних однаково розподілених ВВ дорівнює 1. Оцінити імовірність того, що середнє арифметичне цих ВВ відхиляється від свого математичного сподівання за абсолютною величиною не більше, ніж на 0,1. Порівняти результат із розрахунками за наслідком із центральної граничної теореми.

ПРАКТИЧНЕ ЗАНЯТТЯ №7

1. Система випадкових величин.

2. Закон розподілу двовимірної ДВВ.

3. Функції розподілу двовимірної ВВ. Залежність та незалежність ВВ.

4. Числові характеристики двовимірної ВВ.

5. Функції ВВ та їх характеристики.

Нехай X - НВВ, закон розподілу якої заданий диференціальною функцією (щільністю розподілу імовірностей) $f(x)$, а ВВ $Y = \varphi(X)$. Якщо φ - диференційовна функція, монотонна на усьому проміжку можливих значень X, то щільність розподілу функції $Y = \varphi(X)$ визначається за формулою:

$$g(y) = f[\psi(y)] \cdot |\psi'(y)|, \qquad (*)$$

де $\psi(y)$ - функція, обернена до функції $\varphi(x)$.

Алгоритм знаходження щільності розподілу $g(y)$.

1. Визначити множину можливих значень для $Y = \varphi(X)$.

2. Із функціональної залежності $Y = \varphi(X)$ знайти явний вираз X через Y, тобто функцію $\psi(y)$, обернену до функції $\varphi(x)$.

3. Знайти похідну $\psi'(y)$.

4. За формулою (*) записати щільність розподілу ВВ Y.

5. Перевірити умову нормування для $g(y)$: $\int_{-\infty}^{+\infty} g(y)dy = 1$.

ЗАДАЧІ ДЛЯ РОЗВ'ЯЗУВАННЯ В АУДИТОРІЇ

Приклад 7.1. Знайти закони розподілу компонент двовимірної ВВ, закон розподілу якої заданий таблицею:

X	Y	
	y_1	y_2
x_1	0,01	0,07
x_2	0,25	0,28
x_3	0,12	0,27

Приклад 7.2. ДВВ X задана таблицею

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

Знайти закон розподілу та числові характеристики функції $Y = X^3 - 2$.

Приклад 7.3. ВВ X розподілена за нормальним законом з математичним сподіванням a та середнім квадратичним відхиленням σ. Знайти закон розподілу функції $Y = kX + b$.

Приклад 7.4. ВВ X рівномірно розподілена на [-1;1]. Знайти закон розподілу функції $Y = X^2$.

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Приклад 7.5. Знайти закони розподілу компонент двовимірної ВВ, закон розподілу якої заданий таблицею:

X	Y	
	y_1	y_2
x_1	0,27	0,25
x_2	0,07	0,12

x_3	0,28	0,01

Приклад 7.6. ДВВ X задана таблицею

X	-1	2	5
P	0,3	0,1	0,6

Знайти закон розподілу та числові характеристики функції $Y = X^2 + 1$.

Приклад 7.7. ВВ X розподілена за нормальним законом з математичним сподіванням $a = 2$ та середнім квадратичним відхиленням $\sigma = 3$. Знайти закон розподілу функції $Y = X^5$.

Приклад 7.8. ВВ X рівномірно розподілена на [1;3]. Знайти закон розподілу функції $Y = X^2$.

ПРАКТИЧНЕ ЗАНЯТТЯ №8

1. Статистичні сукупності (генеральна та вибіркова), ознаки та їх розподіли. Числові характеристики статистичних розподілів.

2. Точкові та інтервальні оцінки параметрів статистичних розподілів, вимоги до цих оцінок.

3. Основні формули інтервального оцінювання. Три типи задач вибіркового методу.

Теорема. Імовірність того, що модуль відхилення вибіркової середньої (або частки) від генеральної середньої (або частки) не перевищить число $\Delta > 0$ дорівнює:

$\gamma = P\left(\left|X_B - X_Г\right| \le \Delta\right) = 2\Phi(t)$ (або $\gamma = P\left(\left|w - p\right| \le \Delta\right) = 2\Phi(t)$),

де $\Phi(t)$ - інтегральна функція Лапласа, $t = \dfrac{\Delta}{\mu}$, а μ - **середньоквадратична похибка (стандарт) вибірки**, яка може бути знайдена за наступними формулами:

при оцінюванні середньої кількісної ознаки

$$\mu = \sqrt{\frac{\sigma_Г^2}{n}} \approx \sqrt{\frac{\sigma_B^2}{n}}$$ у випадку повторної вибірки,

$$\mu = \sqrt{\frac{\sigma_Г^2}{n}\left(1-\frac{n}{N}\right)} \approx \sqrt{\frac{\sigma_B^2}{n}\left(1-\frac{n}{N}\right)}$$ у випадку безповторної вибірки;

при оцінюванні частки якісної ознаки

$$\mu = \sqrt{\frac{pq}{n}} \approx \sqrt{\frac{w(1-w)}{n}}$$ у випадку повторної вибірки,

$$\mu = \sqrt{\frac{pq}{n}\left(1-\frac{n}{N}\right)} \approx \sqrt{\frac{w(1-w)}{n}\left(1-\frac{n}{N}\right)}$$ у випадку безповторної вибірки.

Зауваження. При визначенні середньоквадратичної похибки вибірки для частки якісної ознаки буває, що невідомі ні генеральна частка p, ні її точкова оцінка – вибіркова частка w. Тоді добуток pq покладають рівним максимальному можливому значенню - $0{,}25$.

Наслідок. При заданій надійності (довірчій імовірності) $\gamma = 2\Phi(t)$ гранична похибка вибірки дорівнює t-кратній величині стандарту, тобто

$$\Delta = t\mu.$$

Наслідок. Довірчі інтервали (інтервальні оцінки) для генеральної середньої та генеральної частки визначаються формулами:

$$X_B - \Delta \le X_Г \le X_B + \Delta$$

та

$$w - \Delta \le p \le w + \Delta.$$

Із класичних оцінок, в яких точність оцінки визначається граничною похибкою $\Delta = t\mu = \frac{t\sigma_B}{\sqrt{n}}$, можна зробити наступні висновки:

1) при зростанні об’єму вибірки n гранична похибка Δ зменшується, тому точність оцінки збільшується (оскільки звужується довірчий інтервал);
2) збільшення надійності $\gamma = 2\Phi(t)$ оцінки призводить до зростання t (оскільки $\Phi(t)$ - зростаюча функція), а внаслідок цього, і до збільшення Δ. Іншими словами: збільшення надійності класичної оцінки призводить до зменшення її точності;
3) аналіз формул для обчислення середньоквадратичної похибки вибірки μ дозволяє при об’ємах генеральної сукупності $N \to \infty$, або у випадках,

коли $N >> n$ (об'єм генеральної сукупності набагато більший об'єма вибірки), застосовувати більш прості формули повторної вибірки.

ТРИ ТИПИ ЗАДАЧ ВИБІРКОВОГО МЕТОДА.

1) Для заданих об'ємові вибірки та довірчому інтервалі знайти надійність оцінки (дано n і Δ, визначається $\gamma = 2\Phi(t) = P$.
2) При заданих об'ємові вибірки та надійності оцінки знайти довірчий інтервал (дано n і $\gamma = 2\Phi(t) = P$, визначається Δ та $X_B - \Delta \le X_Г \le X_B + \Delta$ або $w - \Delta \le p \le w + \Delta$).
3) При заданих надійності оцінки та довірчому інтервалі знайти необхідний об'єм вибірки (дано $\gamma = 2\Phi(t) = P$ і Δ, знаходиться n).

ЗАДАЧІ ДЛЯ РОЗВ'ЯЗУВАННЯ В АУДИТОРІЇ

Приклад 8.1. В торзі працюють 500 продавців. Серед 100 продавців відібраних за методом безповторної вибірки середній денний виторг склав 2000 грн., а середнє квадратичне відхилення – 40 грн. Знайти імовірність того, що середній денний виторг одного продавця в торзі відрізняється від 2000 грн. не більше, ніж на 10 грн.

Приклад 8.2. Комерційний банк для вивчення можливостей надання довготермінових кредитів населенню провів опитування 1000 чоловік з 10000 своїх клієнтів. Середнє значення необхідного кредиту в вибірці склало 2000 грн., а дисперсія – 1024. Знайти межі довірчого інтервалу для середнього значення кредиту для всіх клієнтів банку з надійністю 0,95.

Приклад 8.3. Вибіркові дослідження показали, що частка покупців, що віддають перевагу новій модифікації товару А, складає 60% від загального числа покупців даного товару. Яким повинен бути обсяг повторної вибірки, щоб з імовірністю 0,9 можна було стверджувати, що частка таких покупців в загальній кількості буде відрізнятися від 0,6 не більше, ніж на 0,05?

Приклад 8.4. Продукція, що вироблена станком-автоматом за зміну, перевіряється методом повторної вибірки. Серед відібраних 400 деталей виявилось 120 першосортних. Знайти імовірність того, що частка першосортних деталей серед усіх вироблених буде відрізнятись від частки таких деталей у вибірці не більше, ніж на 5 % .

Приклад 8.5. Для визначення ефективності внесення добрив було проведено вибіркове обстеження 30 га посівної площі. З кожного гектара відібрали по 1 кв.м. і визначили урожайність на кожному гектарі. Середня урожайність серед обстежених 30 га виявилась 43 ц/га, а дисперсія – 5. В яких межах знаходиться середня урожайність на всій площі, якщо результат необхідно гарантувати з надійністю 0,9.

Приклад 8.6. Коробки з цукерками пакуються автоматично. Середня вага коробки 0,6 кг. На контроль надійшло 2000 коробок. Скільки коробок слід

перевірити методом безповторної вибірки, щоб з ймовірністю 0,9 можна було стверджувати, що середня вага всіх коробок знаходиться в межах від 0,55 до 0,65 кг? Дисперсія ваги не перевищує 0,1.

ЗАДАЧІ ДЛЯ САМОСТІЙНОГО РОЗВ'ЯЗУВАННЯ

Приклад 8.7. Для оцінки частки безробітних серед 5000 робітників одного з районів міста відібрано методом безповторної вибірки 500 чоловік. Виявилось, що в вибірці 25 безробітних. Знайти з надійністю 0,95 довірчий інтервал частки безробітних для всіх робітників району.

Приклад 8.8. В АП 6000 овець. Вибірковим методом було встановлено, що середній настриг вовни з однієї вівці у партії в 1000 голів становить 5 кг, дисперсія 0,9. Знайти ймовірність, з якою можна стверджувати, що середній настриг вовни з однієї вівці для всієї отари відрізнятиметься від 5 кг не більш ніж на 0,2 кг в ту чи іншу сторону.

Приклад 8.9. Вибірковим обстеженням потрібно визначити середню вагу зерна пшениці. Скільки потрібно обстежити зернин, щоб з надійністю 0,9 можна було стверджувати, що середня вага зернини серед відібраних відрізнятиметься від середньої ваги зернини в усій партії не більше, ніж на 0,001 г ? Встановлено, що середнє квадратичне відхилення ваги зернини не перевищує 0,04 г.

Приклад 8.10. Середній вміст вітаміну С серед 100 драже, що перевірялись методом повторної вибірки, склав 14 % . Знайти імовірність того, що середній вміст вітаміну С в усій партії драже буде в межах від 13 % до 15 % , якщо дисперсія ознаки не перевищує 25.

Приклад 8.11. Шляхом безповторної вибірки перевірена якість 1000 деталей з партії в 5000 штук. Серед перевірених було 3 % нестандартних. Визначити межі, в яких знаходиться частка нестандартних деталей в усій партії, якщо результат необхідно гарантувати з ймовірністю 0,9973.

Приклад 8.12. Для оцінки частки деталей найвищого ґатунку в партії з 6000 деталей проводиться вибіркове обстеження. Яким повинен бути обсяг безповторної вибірки, щоб із ймовірністю 0,89 можна було стверджувати, що похибка вибірки не перевищить 0,02?

Індивідуальні науково-дослідні завдання

Зразок ІНДЗ №1

1. Інвестор купив акції двох компаній. Імовірність отримати високі дивіденди на акції першої компанії експерти оцінюють на рівні 0,4 , а другої – 0,6. Знайти імовірність того, що інвестор отримає високі дивіденди по акціям: 1) тільки однієї компанії; 2) хоча б однієї з компаній.

2. Дві незалежні випадкові величини задані своїми законами розподілу:

X	2	5
P	0,6	0,4

У	1	4
P	0,1	0,9

1) Скласти закон розподілу випадкової величини X + 3У.
2) Обчислити $M(X)$, $M(Y)$, $M(3X-2Y-4)$.
3) Обчислити $D(X)$, $D(Y)$, $D(3X-2Y-4)$.

3. Ймовірність виходу з ладу телевізора на протязі гарантійного терміну дорівнює 0,2. За деякий час продано 400 телевізорів. Необхідно знайти:
   1) найімовірніше число телевізорів, що не вийдуть з ладу на протязі терміну гарантії, та імовірність цього числа ;
   2) число телевізорів, що вийдуть з ладу відрізнятиметься від їх середнього числа не більш ніж на 16 в ту чи іншу сторону.

4. Випадкова величина X розподілена за показниковим законом із математичним сподіванням, рівним 2. Знайти :

   1) інтегральну $F(x)$ та диференціальну $f(x)$ функції ВВ X, схематично побудувати графіки $f(x)$ та $F(x)$;

   2) числові характеристики ВВ X та $P(-2<X<3)$.

Зразок ІНДЗ №2

1. (3 бали) Із партії в 7000 деталей перевірена якість 1000 деталей методом безповторної вибірки. Серед відібраних виявилось 95% першого ґатунку. Знайти надійність, з якою можна стверджувати, що частка деталей першого ґатунку в усій партії відрізнятиметься від цієї частки в вибірці не більш ніж на 0,04.

2. (3 бали) Із 5000 робітників підприємства методом безповторної вибірки обстежили зарплату 1000 робітників. Середня вибіркова зарплата виявилась рівною 280 грн., а дисперсія - 640. Визначити з надійністю 0,954 граничні значення середньої зарплати серед усіх робітників підприємства.

3. (4 бали) Необхідно вибірковим методом визначити середню тривалість горіння електролампи в партії із 6000 шт. Встановлено, що середнє квадратичне відхилення не перевищує 100 годин. Похибка вибірки не повинна перевищувати 30 годин. Скільки ламп потрібно відібрати, щоб результат можна було гарантувати з надійністю 0,95? Задачу розв'язати, вважаючи вибірку: а) повторною; б) безповторною.

ІНДЗ №3

Задача1. Із великої партії банкоматів було зроблено вибірку для дослідження закону розподілу часу безперервної роботи банкомату. Результати дослідів наведені в таблиці (N = номеру варіанта) :

Час безвідмовної роботи (год)	Кількість банкоматів	Час безвідмовної роботи (год)	Кількість банкоматів
(1000+N) - (1010+N)	165	(1050+N) - (1060+N)	20
(1010+N) - (1020+N)	120	(1060+N) - (1070+N)	15
(1020+N) - (1030+N)	75	(1070+N) - (1080+N)	10
(1030+N) - (1040+N)	55	(1080+N) - (1090+N)	5
(1040+N) - (1050+N)	35		

Необхідно:

1) побудувати полігон відносних частот (часток), знайти числові характеристики розподілу;

2) обгрунтовано вибрати закон розподілу часу безперервної роботи у генеральній сукупності (теоретичний розподіл), знайти його параметри, функції розподілу та побудувати графік щільності розподілу (на діаграмі для полігону);

3) перевірити за допомогою критерія Пірсона узгодженість вибіркових даних з побудованим теоретичним розподілом (рівень значущості прийняти рівним 0,05).

Використати при розв'язуванні Excel.

Підготуватись до співбесіди.

Задача 2. Для дослідження закону розподілу розміру щомісячної батьківської матеріальної підтримки студентам ОНЕУ проведено вибіркове обстеження, результати якого наведені в таблиці (N – номер варіанта) :

Розмір підтримки($)	Число студ	Розмір підтримки($)	Число студ	Розмір підтримки($)	Число студ
(140+N)-(142+N)	1	(150+N)-(152+N)	114	(160+N)-(162+N)	64
(142+N)-(144+N)	3	(152+N)-(154+N)	186	(162+N)-(164+N)	28
(144+N)-(146+N)	7	(154+N)-(156+N)	200	(164+N)-(166+N)	9
(146+N)-(148+N)	26	(156+N)-(158+N)	172	(166+N)-(168+N)	3
(148+N)-(150+N)	66	(158+N)-(160+N)	120	(168+N)-(170+N)	1

Необхідно:

1) побудувати полігон відносних частот (часток), знайти числові характеристики розподілу;

2) обгрунтовано вибрати закон розподілу розміру щомісячної батьківської матеріальної підтримки у генеральній сукупності (теоретичний розподіл), знайти його параметри, функції розподілу та побудувати графік щільності розподілу (на діаграмі для полігону);

3) перевірити за допомогою критерія Пірсона узгодженість вибіркових даних з побудованим теоретичним розподілом (рівень значущості прийняти рівним 0,05).

Використати при розв’язуванні Excel.

Підготуватись до співбесіди.

Задача 3. Проведені спостереження, в результаті яких реєструвалась кількість покупців за проміжок часу. Вважаючи, що течія найпростіша (Пуассонівська), визначити її параметри та перевірити за критерієм Пірсона узгодженість побудованого теоретичного закона реальним змінам вхідної течії. На одній діаграмі побудувати полігони емпіричних та теоретичних частот. Дані наведені у таблиці:

Варіант	Показники розподілу	Результати спостережень
1,11,21	Кількість покупців за 5 хв.	2 3 4 5 6 7 8 9 10
	Кількість (частоти) інтервалів	1 2 3 3 3 3 2 2 2
2,12,22	Кількість покупців за 5 хв.	0 1 2 3 4 5 6 7
	Кількість (частоти) інтервалів	2 3 6 7 6 4 2 1
3,13,23	Кількість покупців за 5 хв.	4 5 6 7 8 9 10 11 12
	Кількість (частоти) інтервалів	2 3 4 4 4 4 3 2 2
4,14,24	Кількість покупців за 5 хв.	1 2 3 4 5 6 7
	Кількість (частоти) інтервалів	5 6 6 6 4 2 1

5,15,25	Кількість покупців за 5 хв. Кількість (частоти) інтервалів	3 4 5 6 7 8 9 10 4 5 6 6 5 4 3 2
6,16,26	Кількість покупців за 10 хв. Кількість (частоти) інтервалів	3 4 5 6 7 8 9 10 3 4 5 5 4 3 2 2
7,17,27	Кількість покупців за 10 хв. Кількість (частоти) інтервалів	0 1 2 3 4 5 6 7 8 9 1 1 3 6 7 7 6 5 3 1
8,18,28	Кількість покупців за 10 хв. Кількість (частоти) інтервалів	1 2 3 4 5 6 7 4 6 7 6 5 4 3
9,19,29	Кількість покупців за 10 хв. Кількість (частоти) інтервалів	4 5 6 7 8 9 10 11 12 13 14 2 3 3 4 4 4 4 4 3 3 3
10,20, 30	Кількість покупців за 10 хв. Кількість (частоти) інтервалів	7 8 9 10 11 12 13 14 15 16 1 3 3 4 5 7 5 4 3 3

Використати при розв'язуванні Excel.

Підготуватись до співбесіди.

ІНДЗ №4

Задача. Залежність між випуском продукції **У** (тон) протягом доби та величиною основних виробничих фондів (ОВФ) **Х** (млн.грн.) для сукупності 50 однотипних підприємств наведена в таблиці (N – номер варіанта):

У / **Х**		7+N– 11+N	11+N– 15+N	15+N- 19+N	19+N- 23+N	23+N- 27+N	m_x
20+N-	25+N	2	1				3
25+N-	30+N	3	6	4			13
30+N-	35+N		3	11	7		21
35+N-	40+N		1	2	6	2	11
40+N-	45+N				1	1	2
m_y		5	11	17	14	3	50

Необхідно:

1) побудувати точкову діаграму статистичної залежності (кореляційне поле); визначити аргументи (регресори), які впливають на функцію-регресант;

2) побудувати моделі регресійної залежності **У** на **Х** та **Х** на **У**. Оцінити щільність кореляційного зв'язку;

3) використати моделі для економічного аналізу та прогнозування.

Використати при розв’язуванні Excel.

Підготуватись до співбесіди.

Література.

1. Барковський В.В., Барковська Н.В., Лопатін О.К. Математика для економістів. Теорія ймовірностей та математична статистика. – К.: Національна академія управління, 2005.
2. Чернышев В.Г. Теория вероятностей и математическая статистика. – Одесса: Оптимум, 2004.
3. Мур Джеффри, Уэдерфорд Ларри Р., Эппен Г., Гулд Ф., Шмидт Ч. Экономическое моделирование в Microsoft Exel. – М.: Вильямс, 2004.
4. Кремер Н.Ш. Теория вероятностей и математическая статистика.
– М.: ЮНИТИ – ДАНА, 2002.

5. *Мацкул В.М.* Теорія ймовірностей та математична статистика: навчальний посібник для студентів ОНЕУ денної форми навчання усіх спеціальностей.- Одеса: ОНЕУ, 2017.