Статистичні ряди розподілу

- Статистичний ряд розподілу — впорядкований розподіл одиниць досліджуваної сукупності на групи за групувальною (варіативною) ознакою. Вони характеризують склад (структуру) досліджуваного явища, дозволяють судити про однорідність сукупності, межі її зміни, закономірності розвитку досліджуваного об'єкта.

- *Ряди розподілу одиниць сукупності за ознаками, що мають кількісний вираз, називаються **варіаційними рядами.** У таких рядах значення ознаки (варіанти) знаходяться в порядку зростання чи спадання.*
- *У варіаційному ряді розподілу розрізняють два елементи: варіанта і частота. **Варіанта** - це окреме значення групувальної ознаки, **частота** - число, яке показує, скільки разів зустрічається кожна варіанта.*
- *У математичній статистиці обчислюється ще один елемент варіаційного ряду - **частість.** Остання визначається, як відношення частоти випадків даного інтервалу до загальної суми частот. Частість визначається в частках одиниці, відсотках (%) в проміле (%о).*
- *Таким чином, варіаційний ряд розподілу - це такий ряд, у якому варіанти розташовані в порядку зростання або спадання, вказані їх частоти або частості. Варіаційні ряди бувають дискретні (переривні) і інтервальні (неперервні).*
- ***Дискретні варіаційні ряди** - це такі ряди розподілу, в яких варіанта як величина кількісної ознаки може приймати тільки певне значення. Варіанти різняться між собою на одну чи кілька одиниць.*
- *Так, кількість вироблених деталей за зміну конкретним робітником може виражатися тільки одним певним числом (6, 10, 12 і т.д.). Прикладом дискретного варіаційного ряду може бути розподіл працівників за кількістю вироблених деталей.*

Дискретний ряд розподілу

Вироблено деталей за зміну, шт. (x>)	*Кількість робітників, ЧОЛ.,()*
6	*16*
7	*10*
8	*8*
9	*10*
10	*12*
11	*16*
12	*3*

Інтервальний ряд розподілу _

- ***Інтервальні (неперервні) варіаційні ряди*** *- такі ряди розподілу, в яких значення варіанти дано у вигляді інтервалів, тобто значення ознак можуть відрізнятися одне від одного на скільки завгодно малу величину. При побудові варіаційного ряду неперервної ознаки неможливо вказати кожне значення варіанти, тому сукупність розподіляється за інтервалами. Останні можуть бути рівні і нерівні. Для кожного з них вказуються частоти або частості (табл. 19).*
- *В інтервальних рядах розподілу з нерівними інтервалами обчислюють такі математичні характеристики, як щільність розподілу і відносна щільність розподілу на даному інтервалі. Перша характеристика визначається відношенням частоти до величини того ж інтервалу, друга - відношенням частості до величини того ж інтервалу. Для наведеного вище прикладу щільність розподілу на першому інтервалі становитиме 3:5 = 0,6, а відносна щільність на цьому інтервалі - 7,5:5= 1,5%.*

Інтервальний ряд розподілу _

Чисельність працюючих, чол. (x)	*Кількість цехів (",)*	*% до підсумку*
20-25	*3*	*7,5*
25-30	*9*	*22,5*
30-35	*16*	*40,0*
35-40	*8*	*20,0*
40-45	*4*	*10,0*
Всього	*40*	*100,0*

Графічне зображення рядів розподілу.

- *Графічне зображення рядів розподілу (як і статистичних даних взагалі), крім досягнення наочності, переслідує й аналітичну мету. Графік дозволяє в найбільш простій і доступній формі піддати аналізу (візуально) статистичний ряд розподілу. Варіаційні ряди залежно від виду і поставленої задачі їх аналізу графічно можуть бути зображені у виді полігону, гістограми, кумуляти, огіви.*

Розподіл робітників за змінною виробіткою продукції

Вироблено деталей за зміну, шт. (x;)	*Кількість робітників, чол. (п)*
30	*101*
31	*125*
32	*250*
33	*306*
34	*150*
35	*86*
Разом	*1018*

• ***Полігон*** *розподілу будується в прямокутній системі координат, при цьому на осі абсцис відкладається варіанта, а на осі ординат -частота або частість. За допомогою полігону розподілу, як правило, графічно зображуються дискретні варіаційні ряди (рис. 3, табл. 20). При побудові полігону для інтервальних рядів розподілу ордината, яка відповідає частоті (частості) встановлюється перпендикулярно осі абсцис у точці, що відповідає центру інтервалу.*

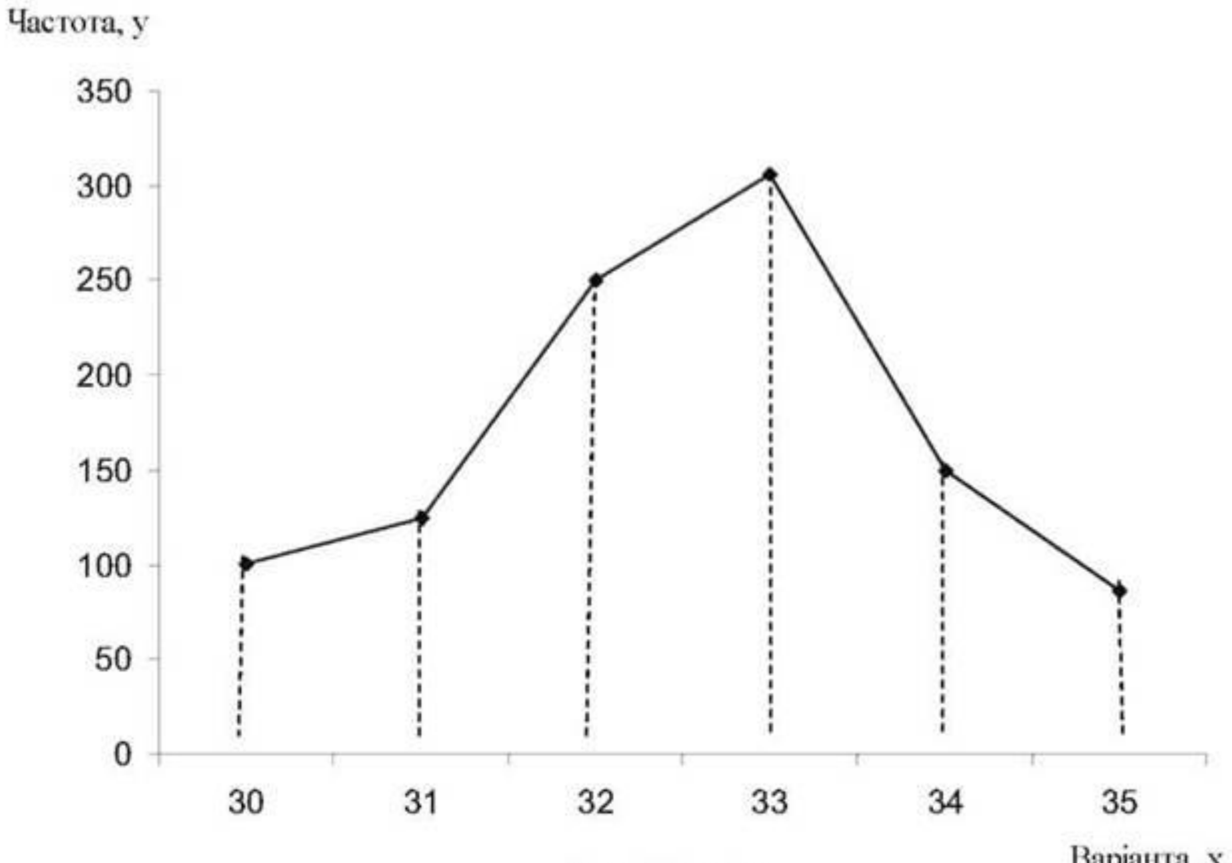


Рис.3 Полігон

Такий спосіб зображення інтервального ряду називають способом "навантажених ординат". Він заснований на допущенні умови рівномірного розподілу частот у межах інтервалів. Якщо це так, то частоти можна віднести до конкретного значення варіанти, яке знаходиться в центрі інтервалу. За таких умов середина (центр) інтервалу ніби "навантажується". Для графічного зображення інтервальних (неперервних) варіаційних рядів частіше використовуються гістограми. Остання являє собою ступінчасту фігуру у вигляді прямокутників, що примикають один до одного. Порядок побудови цього виду графіків такий. На осі абсцис відкладають інтервали значень варіанти. Вони ж є основами прямокутників, висота яких (ордината) пропорційна частоті (частості) інтервалів

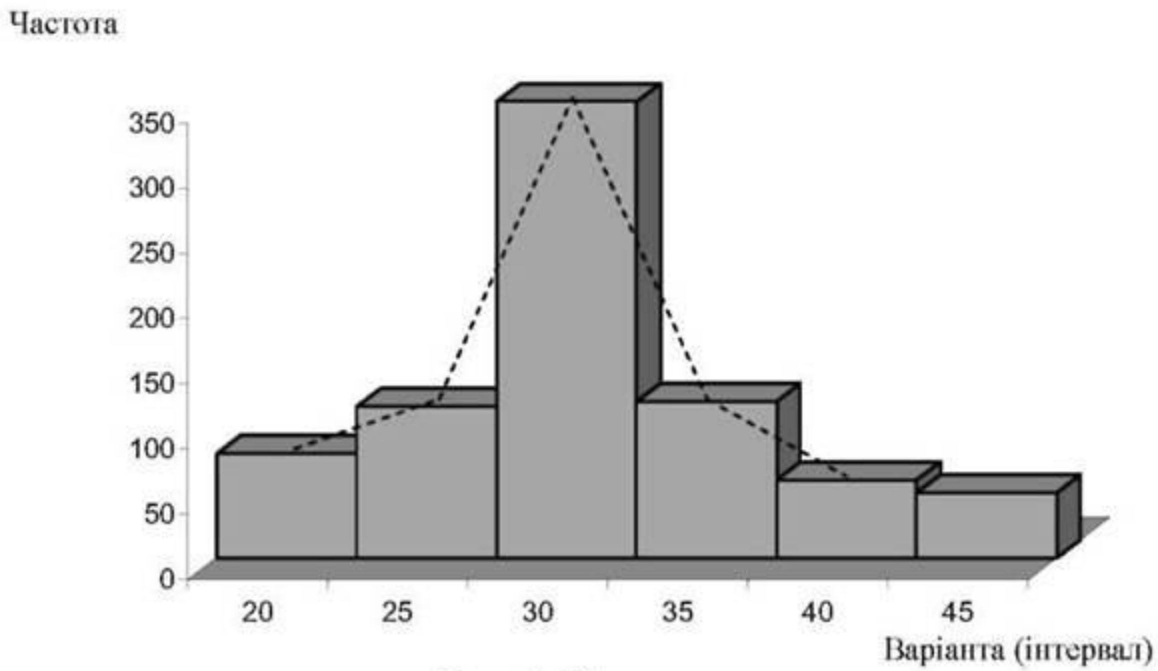


Рис. 4. Гістограма

Таблиця 21

Розподіл робітників за змінною виробіткою продукції

Вироблено деталей за зміну, шт.(xi)	*Кількість робітників, чол. (п)*
20-25	*80*
25-30	*116*
30-35	*350*
35-40	*120*
40-45	*60*
45-50	*50*
Всього	*776*

- *Якщо необхідно зобразити на графіку інтервальний ряд розподілу з нерівними інтервалами, то гістограму будують не по частотах (частостях) інтервалів, а по показниках щільностей розподілу. При побудові гістограми по абсолютній щільності розподілу загальна площа її дорівнюватиме чисельності сукупності. При побудові графіка відносної щільності площа гістограми дорівнюватиме одиниці.*
- *Інколи ідуть шляхом розчленування (подрібнення) інтервалів варіаційного ряду. У цьому випадку на графіку площа гістограми (дрібно ступінчастої) повинна відповідати попередній величині. Якщо процес подрібнення інтервалів продовжити, одержуючи все дрібніші їх параметри, то дрібноступінчаста гістограма в межі являтиме плавну криву.*
- *При зображені варіаційного ряду з нагромадженими частотами (частостями) у прямокутні системі координат одержується так звана крива сум - кумулята. Якщо розподіл носить дискретний характер (переривний) на графіку на осі абсцис відкладають значення варіанти, на осі ординат - накопичені частоти (частості) (рис. 5, табл. 22).*

Таблиця 22
Розподіл підприємств за рівнем рентабельності

Рівень рентабельності, % (хО	Кількість підприємств, (п)	Накопичена частота (п, ')
25	3	3
26	6	9
27	10	19
28	8	27
29	6	33
30	2	35
Всього	35	X

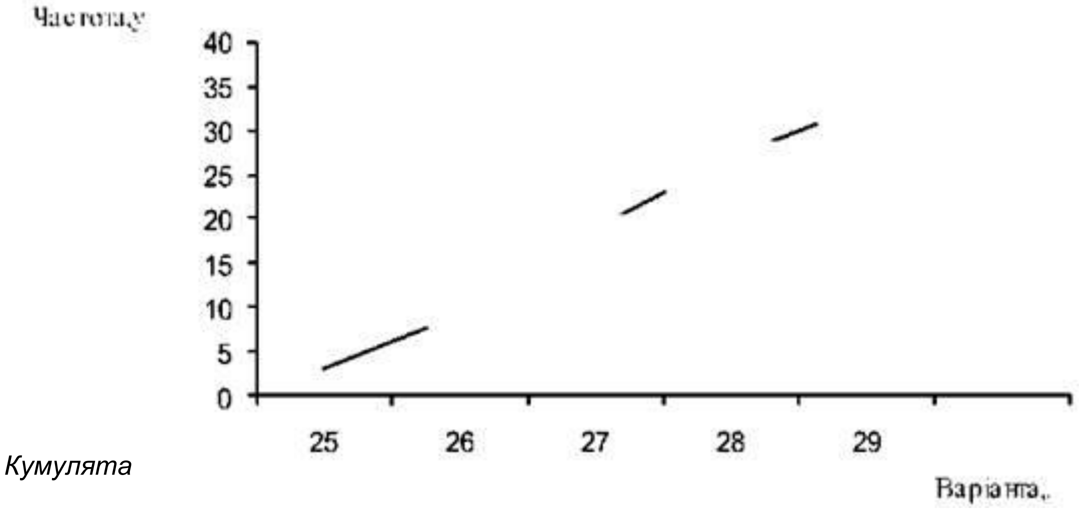
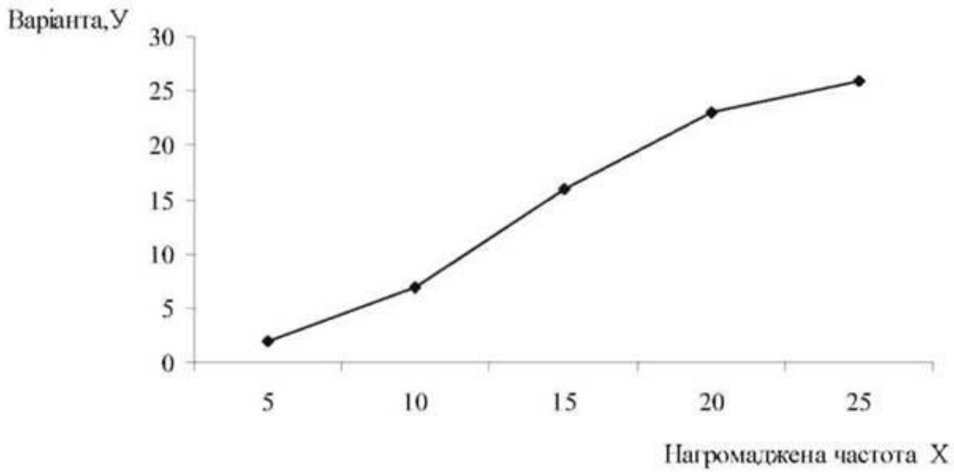


Рис. 5. Кумулята

- *При побудові кумуляти для інтервального ряду розподілу будують точки, абсциси яких - праві границі інтервалів, а ординати - відповідні їм нагромаджені частоти або частості.*
- *По інтервальному (неперервному) ряду розподілу будують точки, абсциси яких - праві границі інтервалів, а ординати - частоти (частости), що їм відповідають.*
- *Аналогічно кумуляті в прямокутній системі координат будують огіву. Різниця графіка лише в тому, що на осі абсцис наносять нагромаджені частоти, а на осі ординат - значення варіант. Щоб побачити форму огіви на графіку кумуляти, досить лист паперу з її зображенням повернути на 90° і подивитися на нього з протилежної сторони (на світло) (рис. 6, табл. 23).*

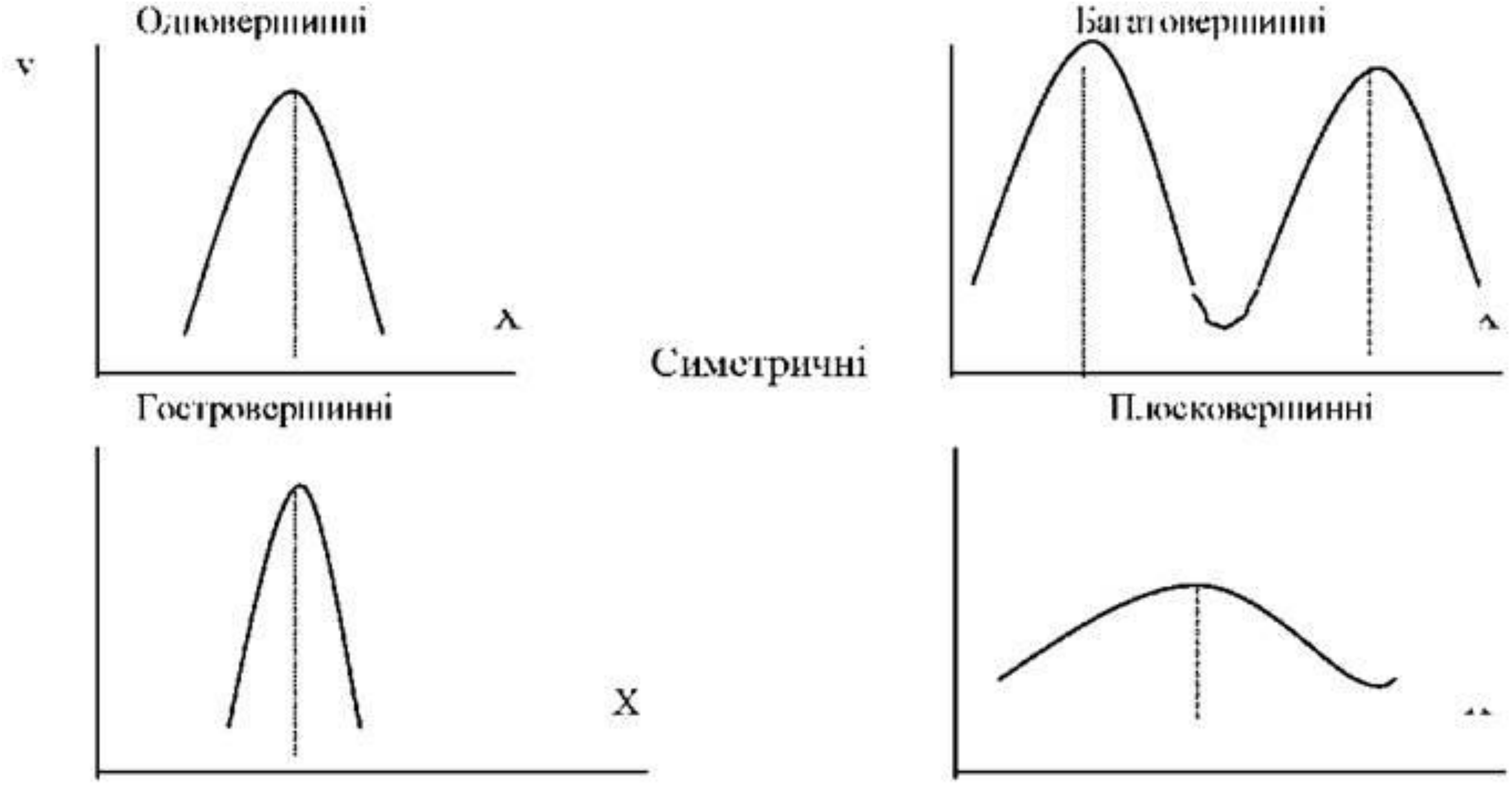
Розподіл підприємств за рівнем рентабельності

*Рівень рентабельності, % **(*0***	*Кількість підприємств (-)*	*Нагромаджена частота (щ ')*
20-23	*2*	*2*
23-26	*5*	*7*
26-29	*9*	*16*
29-32	*7*	*23*
32-35	*3*	*26*
Разом	*26*	*X*



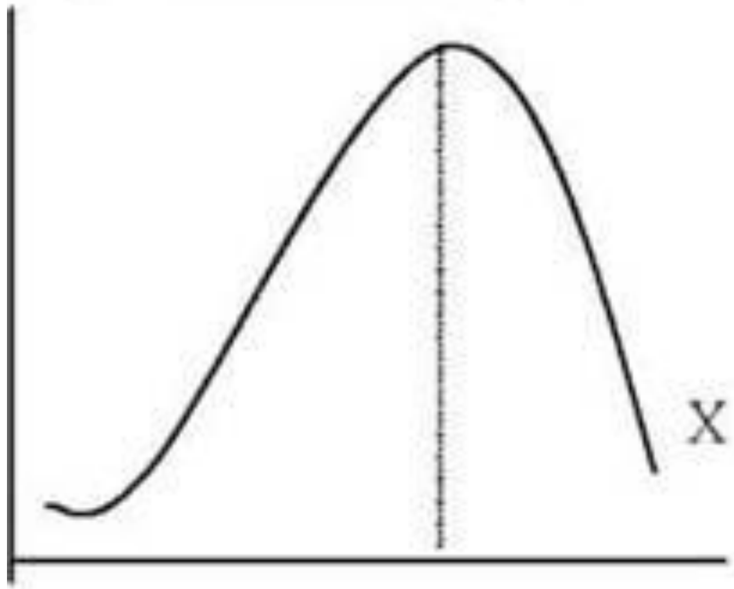
Якщо збільшити до нескінченності число одиниць спостережень і зменшити величину інтервалу, графік ряду розподілу буде мати форму кривої

- Різноманітність статистичних сукупностей – передумова різних форм співвідношення частот і значень ознаки, що варіює.
- За своєю формою ряди розподілу поділяють на такі види: одно-, дво- і багатовершинні. Наявність двох і більше вершин свідчить про неоднорідність сукупності, про поєднання в ній груп з різними рівнями ознаки. Ряди розподілу якісно однорідних сукупностей, як правило, одновершинні.
- Серед одновершинних рядів розподілу є симетричні (скошені), гостро- і плосковершинні.

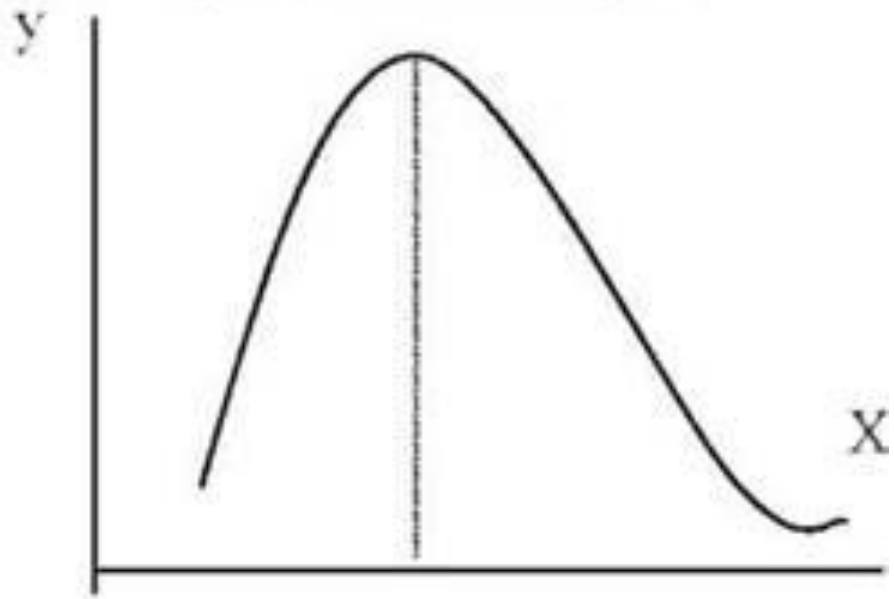

Одновершинні
Багатовершинні
Симетричні
Гостровершинні
Плосковершинні
X

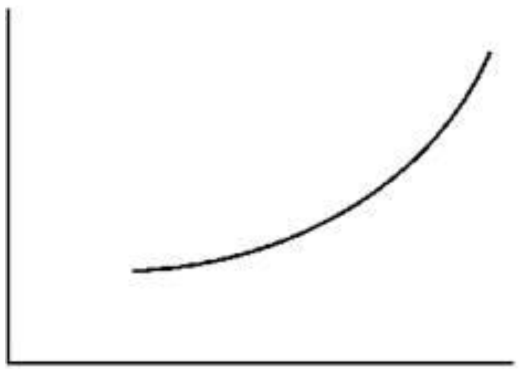
Помірно асиметричні

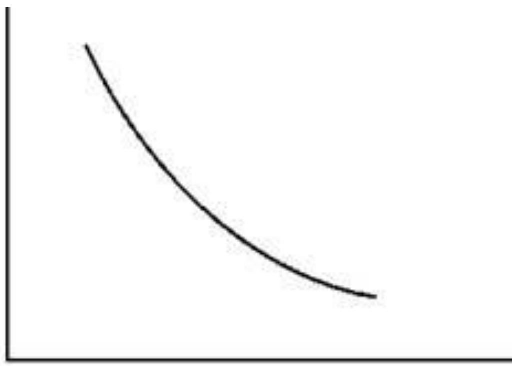
Лівостороння скошеність
(від'ємна асиметрія)



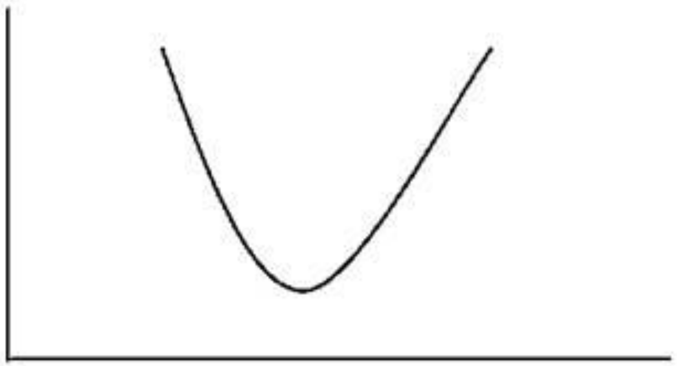
Правостороння скошеність
(додатня асиметрія)

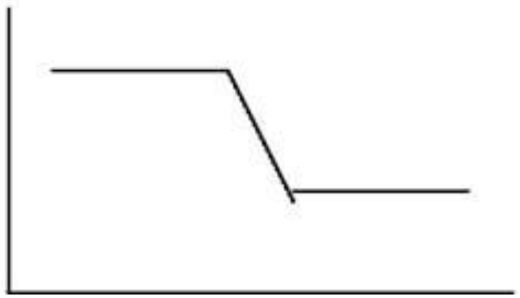


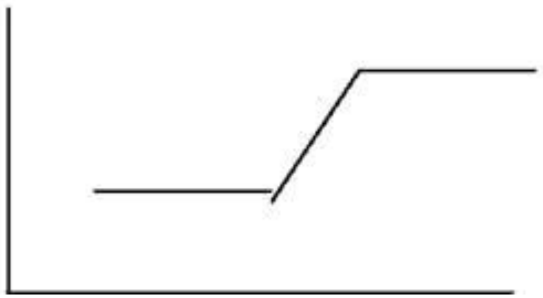




U-поді







Криві розподілу та способи перевірки гіпотез

- **3.** Аналіз варіації рядів розподілу дає змогу розкрити закономірності співвідношення варіантів і частот за допомогою функції теоретичної кривої. Серед найпоширеніших графічних зображень є крива нормального розподілу.
- Її використовують як стандарт для порівняння інших розподілів.
- Математична статистика використовує кілька показників, за допомогою яких можна оцінити, наскільки фактичний розподіл узгоджується з нормальним. Такі показники називають критеріями узгодження.
- Критерії узгодження - це певна величина, що оцінює досліджуване явище з певною ймовірністю.
- Статистика використовує критерії узгодження Пірсона, Колмогорова, Ястремського, Романовського, Фішера та ін.
- Для перевірки гіпотези про відповідність чи невідповідність теоретичного закону розподілу емпіричному можна використовувати будь-який з перелічених критеріїв, які забезпечують дослідження закону розподілу з різною точністю, надійністю і трудомісткістю.

- 3. Серед найпоширеніших теоретичних розподілів найчастіше використовується нормальний розподіл (розподіл Лапласа – Гауса), який відображає нормальна крива (симетрична відносно max ординати). Такий розподіл є результатом впливу на значення ознаки необмеженої кількості незалежних один від одного факторів, як це буває в природі.
- Нормальний розподіл повністю визначений двома параметрами – середньою арифметичною (хсер) та стандартним (середнім квадратичним) відхиленням (σ). Значення ознаки при нормальному розподілі переважно зосереджуються біля центра розподілу – хсер.
- Значення ознаки, які істотно відхиляються від хсер, зустрічаються рідко. Підпорядкованість цьому закону буде більш точнішою у разі одночасної дії великої кількості випадкових величин. Якщо жодна з випадково діючих причин за своєю дією не буде переважаючою над іншими, то закон розподілу дуже близько підходить до нормального. Наприклад, за нормальним законом розподілені вага та зріст людини, відхилення у виробничому процесі при нормальному рівні організації та технології, в розподілі населення означеного віку за розміром взуття та багато інших явищ, у яких проявляється велика кількість незалежних значень спостережуваних ознак, серед яких немає суттєво відмінних від решти значень ознаки статистичної сукупності.
- Поняття нормального розподілу покладено в основу багатьох методів статистики.
- Його використовують як стандарт для порівняння інших розподілів, застосовують у вибірковому, кореляційно-регресивному, факторному методах статистичного дослідження.
- Нормальний розподіл подібний до одновершинних симетричних розподілів, тому його розглядають як перше наближення в разі статистичного моделювання. При розв’язуванні багатьох задач необхідно встановити, за яким законом розподілена ознака статистичної сукупності. Щоб перевірити нормальність розподілу, тобто відповідність досліджуваного розподілу нормальному, слід частоти фактичного розподілу порівняти з теоретичними частотами, характерними для нормального розподілу. Для цього за фактичними даними, підставляючи їх у відповідну формулу, обчислюють теоретичні частоти кривої нормального розподілу. Ступінь відповідності між фактичним і теоретичним розподілами оцінюється за допомогою показників (критеріїв) узгодженості.

Критерій згоди – це статистичний показник, що використовується для з’ясування розбіжностей між прийнятою статистичною моделлю і спостережуваними даними ознаки, які має описати ця модель. Одним з найпоширеніших є критерій Пірсона (хі-квадрат) – χ^2:

$$\chi^2 = \sum (f_{теор.} - f_{емп.})^2 / f_{теор.}$$

Чим більшою буде різниця між емпіричними та теоретичними частотами, тим більше буде значення критерію Пірсона.

Обчисливши фактичне значення критерію, порівнюють його з табличним (критичним) значенням:

якщо $\chi^2_{ф.} > \chi^2_{табл.}$, тобто χ^2 попадає у критичну область – розбіжність між емпіричними та теоретичними частотами є істотною і її не можна пояснити випадковими коливаннями даних, а емпіричний розподіл є принципово відмінним від теоретичного;

якщо $\chi^2_{ф.} < \chi^2_{табл.}$, відхилення фактичних частот від теоретичних вважається випадковим, неістотним; емпіричний розподіл відповідає теоретичному.

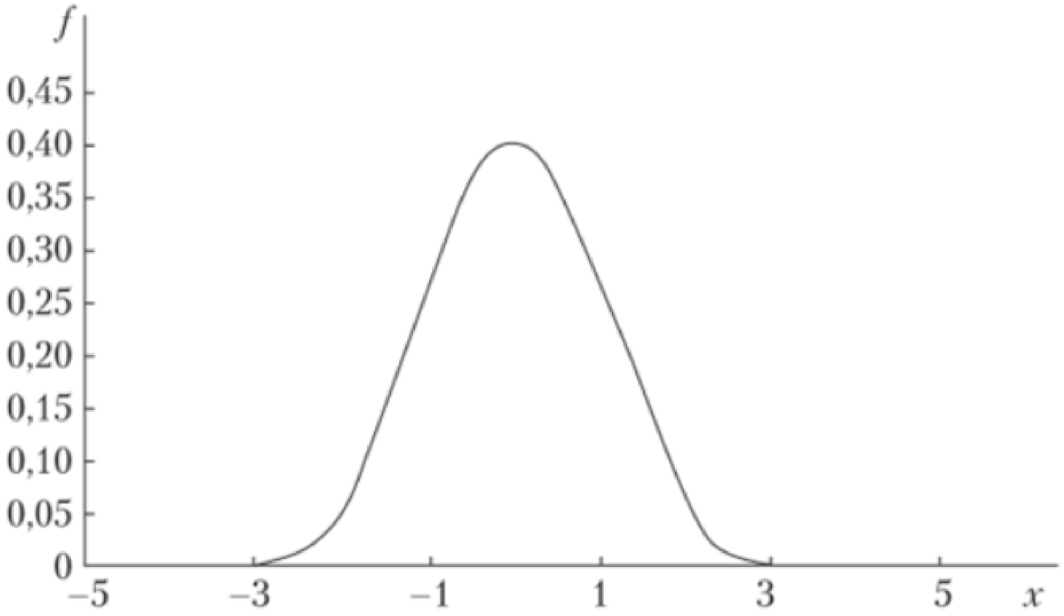
Існують ще такі критерії перевірки відповідності нормальному розподілу як критерій Колмогорова-Смирнова, Романовського, Ястремського, Фішера, Вілконсона. Необхідною умовою використання цих критеріїв є достатньо велике число спостережень – не менше 100 (20-30 за критерієм Пірсона).

Розподіл неперервної випадкової величини *x* називають нормальним N (*x*, a), якщо відповідна їй щільність розподілу виражається формулою

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

Нормальний розподіл ознаки спостерігається в тих випадках, коли на величину варіантів, що входять до складу варіаційного ряду, діє безліч випадкових, незалежних або слабо залежних факторів, кожен з яких відіграє в загальній сумі незначну роль. Порушення нормального характеру розподілу часто є свідченням неоднорідності сукупності.

Крива нормального розподілу являє собою одновершинная симетричну колоколообразную фігуру, права і ліва гілки якої рівномірно і симетрично зменшуються, асимптотично наближаючись до осі абсцис (рис. 5.1). Відмінною особливістю цієї кривої є збіг в ній середньої арифметичної, моди і медіани. Якщо всю площу між кривою нормального розподілу і віссю абсцис прийняти за 100%, то в межах однієї *про* укладено 68,3% частот, в межах двох а - 95,4%, в межах трьох а - 99,7% («правило трьох сигм »).



- Нормальний розподіл має такі властивості.
- 1. Значення ознаки *x – Xст* мають тенденцію концентруватися близько точки *t* = 0, де *t =* -є нормірованим відхиленням.
- 2. Нормальна крива симетрична щодо вертикальної осі.
- 3. Значення спостережень не обмежені по своїй величині.
- 4. *x* , Мо, Ме мають одне і те ж значення при *t* = 0.
- 5. Зміна величини *t* характеризує різні типи розподілу.

РОЗПОДІЛ ВЕЙБУЛЛА

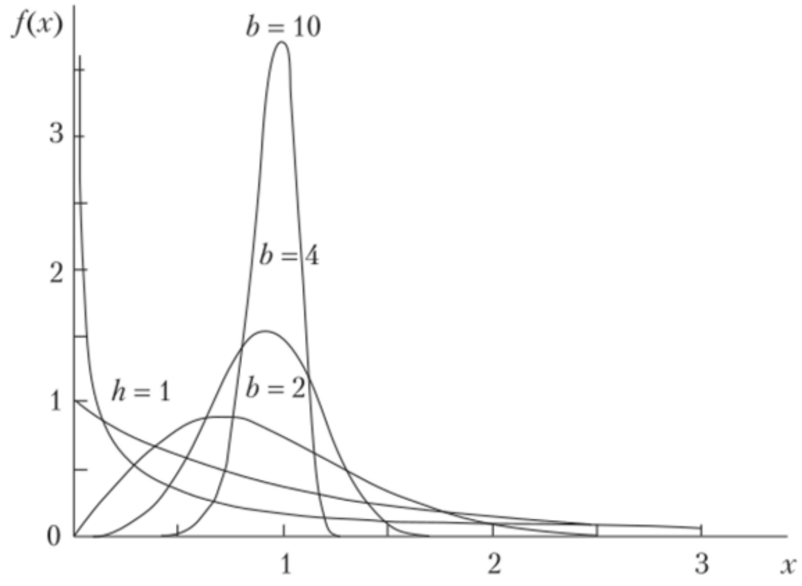
Розподіл Вейбулла названо в честь шведського дослідника Валодді Вейбулла (1887-1979), котрий застосовував цей розподіл для опису часів відмов різного типу в теорії надійності. В даний час розподіл Вейбулла інтенсивно використовується не тільки в теорії надійності, але і в страхуванні.

Щільність розподілу Вейбулла записується у вигляді

$$f(x)=\frac{b}{a}\left(\frac{x}{a}\right)^{b-1}e^{-\left(\frac{x}{a}\right)^{b}},$$

де a - параметр масштабу; b - параметр форми.

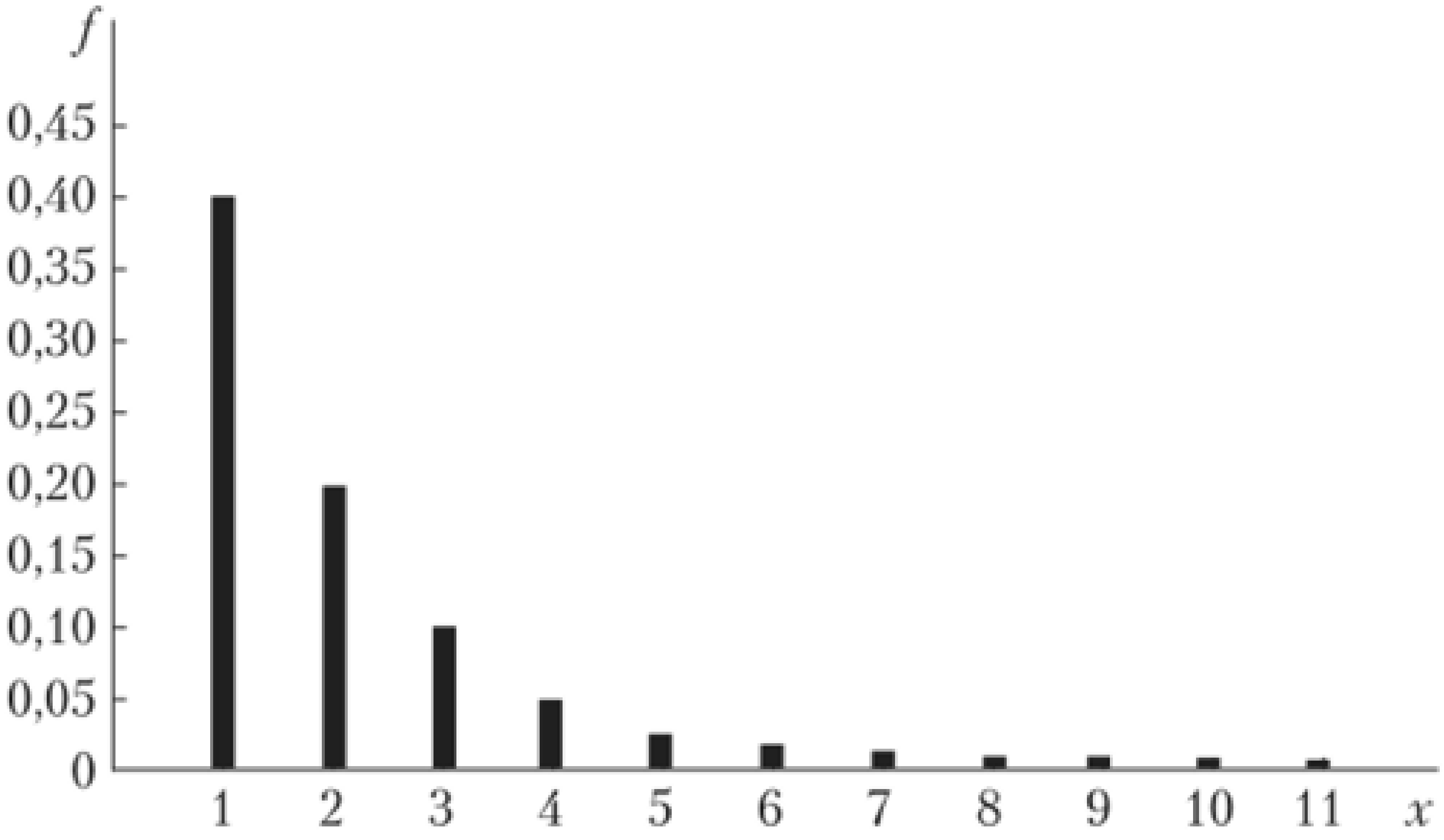
Параметр b визначають за спеціальними таблицями залежно від коефіцієнта варіації. При b = 1 розподіл Вейбулла близько до експоненціального закону розподілу, при h = 2,5-ь 3,5 - до нормального. Внаслідок такого видозміни розподіл Вейбулла вважають дуже гнучким законом.



РОЗПОДІЛ ПУАССОНА

- Розподіл Пуассона називають розподілом рідкісних явищ, воно спостерігається в сумах, число одиниць яких досить велике *(N*> 100), а частка одиниць, що володіють великими значеннями ознаки, мала. Розподіл Пуассона також відноситься до числа найважливіших теоретичних розподілів, що мають практичне застосування. Прикладами змінних, розподілених по закон} 'Пуассона, можуть служити число дефектів у виробничому процесі, число відмов технологічного обладнання і т.д.
- Класичну форму розподіл Пуассона приймає в тому випадку, якщо значення ознаки носять дискретний характер, де в міру збільшення значень ознак частоти різко зменшуються і $x_{ср} = a^2$. Розподіл Пуассона визначається формулою (де *а* - середня арифметична ряду *(х = а).*

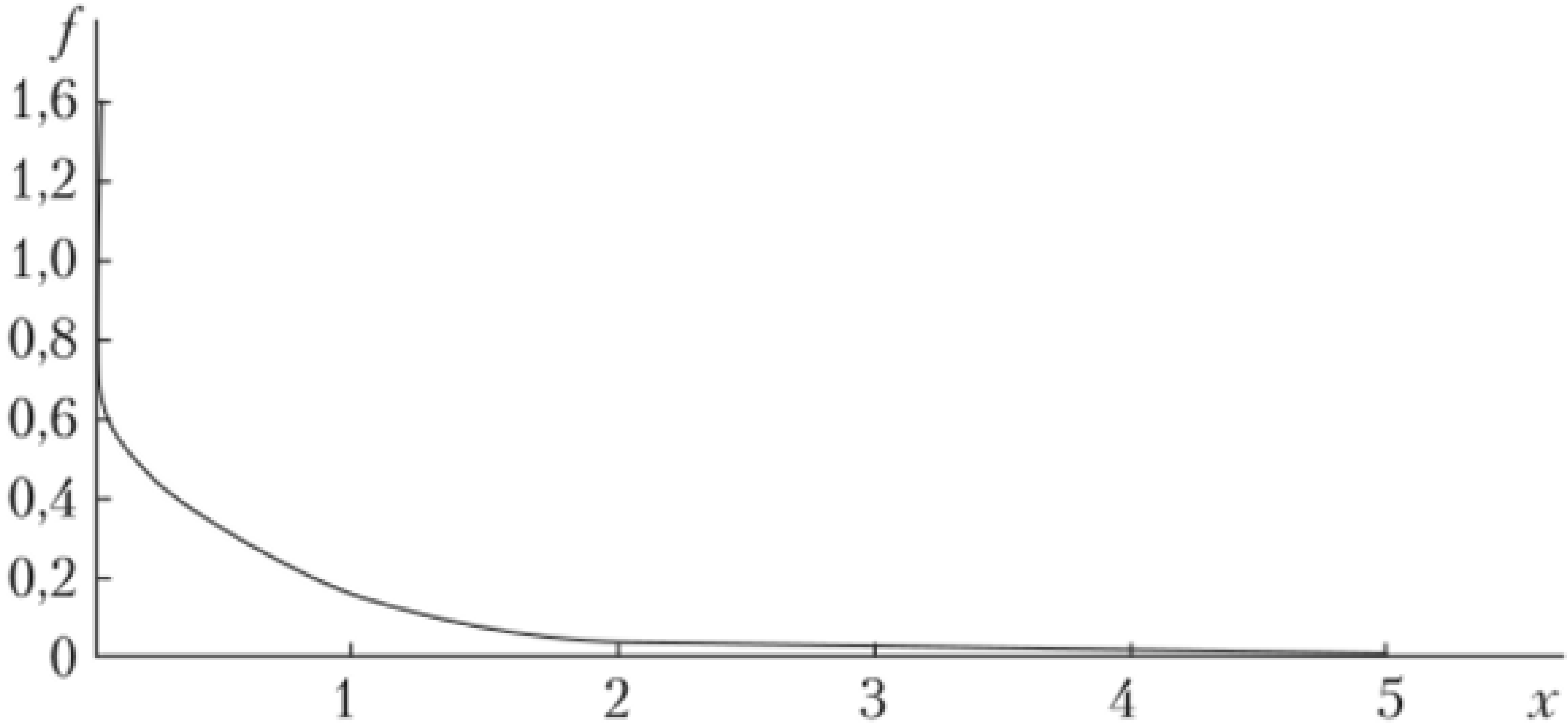
$$f(x) = \frac{a^x e^{-a}}{x!},$$


f
0,45
0,40
0,35
0,30
0,25
0,20
0,15
0,10
0,05
0
1
2
3
4
5
6
7
8
9
10
11
x

ЕКСПОНЕНЦІАЛЬНИЙ РОЗПОДІЛ

- Експонентний закон характерний для розподілу випадкових величин, зміна яких обумовлено впливом якогось домінуючого фактора. Експоненціальне розподіл часто використовується для опису інтервалів між послідовними випадковими подіями, наприклад інтервалів між заходами на непопулярний сайт, так як ці відвідування є рідкісними подіями.
- Показовий розподіл є окремим випадком розподілу Вейбулла. Цей розподіл має тільки один параметр, який і визначає його характеристики. Щільність розподілу має такий вигляд:

$$f(x) = \lambda e^{-\lambda x} ,$$



f
1,6
1,2
1,0
0,8
0,6
0,4
0,2
0
1
2
3
4
5
x