ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ імені ІВАНА ФРАНКА

Факультет прикладної математики та інформатики

Теорiя ймовiрностi та математична статистика

# Iндивiдуальне завдання №1

*Виконав:*

Студент групи ПМі-21

Процьків Назарій

*Викладач:*

Пелюшкевич Ольга

Володимирівна

2023

Постановка задачі

Згенерувати вибірку заданого об’єму (не менше 50) з вказаного проміжку для дискретної статистичної змінної.

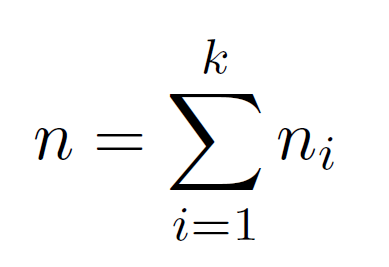
1. На підставі отриманих вибіркових даних: побудувати варіаційний ряд та частотну таблицю; представити графічно статистичний матеріал, задати аналітично та побудувати емпіричну функцію розподілу; обчислити числові характеристики дискретного розподілу.
2. Утворити інтервальний розподіл варіанти, побудувати гістограму розподілу, побудувати емпіричну функцію розподілу для інтервального розподілу, обчислити числові характеристики для згрупованих даних.

Короткі теоретичні відомості

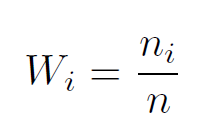
Кiлькiснi ознаки елементiв генеральної сукупностi можуть бути одновимiрними i багатовимiрними, дискретними i неперервними.

Коли реалiзується вибiрка, кiлькiсна ознака, наприклад *X*, набуває конкретних числових значень X = xi якi називають варiантою.

Зростаючий числовий ряд варiант називают варiацiйним. Кожна варiанта вибiрки може бути спостереженою *ni* раз (*ni* >=1), число *ni* називають частотою варiанти *xi*.



Вiдношення частоти *ni* варiанти *xi* називають її вiдносною частотою i позначають через *Wi*, тобто



Множина всiх можливих значень випадкової величини *X* називається генеральною сукупнiстю, а множина значень *xi* (*i* = 1*,* 2*..., k*), яка одержана в результатi випробувань, вибiркою з генеральної сукупностi або статистичною сукупнiстю. Число елеменiтв вибiрки називається обсягом вибiрки.

Послiдовнiсть варiант, записаних за зростанням, називається варiацiйним рядом (дискретним варiацiйним рядом).

Якщо дослiджується ознака генеральної сукупностi *X*, яка є неперервною, то варiант буде багато. У цьому разi варiацiйний ряд - це певна кiлькiсть рiвних або нерiвних частинних iнтервалiв чи груп варiант зi своїми частотами.

Такi частиннi iнтервали варiант, якi розмiщенi у зростаючiй послiдовностi, утворюють iнтервальний варiацiйний ряд.

## Дискретний статистичний розподiл вибiрки та її характеристики

Перелiк варiант варiацiйного ряду i вiдповiдні їм частот, або вiдносних частот, називають дискретним статистичним розподiлом вибiрки.

У табличнiй формi вiн має такий вигляд:

A screenshot of a graph

Description automatically generated with low confidence

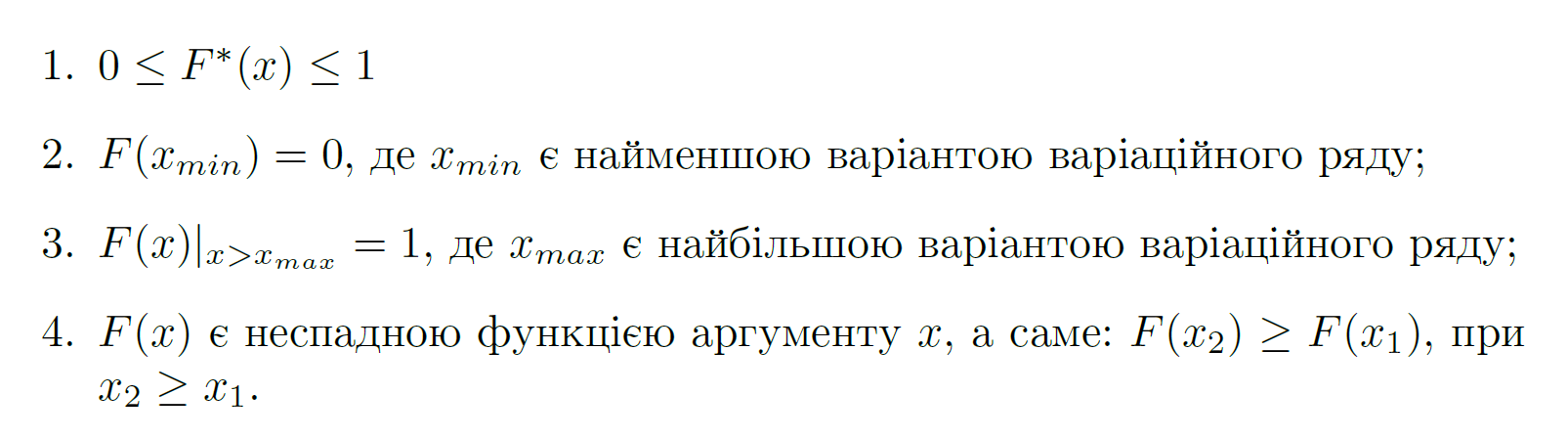
### Емпiрична функцiя розподiлу

### Функцiя аргументу *x*, що визначає вiдносну частоту подiї *X < x*, тобто

### 

називається емпiричною, або комулятою. Тут *n* - обсяг вибiрки; *ni* - кiлькiсть варiант статистичного розподiлу вибiрки, значення яких менше за фiксовану варiанту *x*; *F* ∗(*x*) - називають ще функцiєю нагромадження вiдносних частот.

Властивостi *F* ∗(*x*):

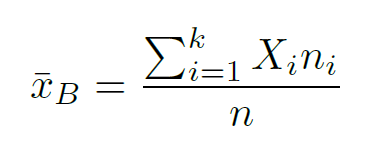
2.3 Полiгон частот i вiдносних частот.

Дискретний статичтичний розподiл вибiрки можна зобразити графiчно у виглядi ламаної лiнiї, вiдрiзки якої сполучають координати точкок (*xi*; *ni*), або (*xi*; *Wi*).

У першому випадку ламану лiнiю називають полiгоном частот, у другому - полiгоном вiдносний частот.

2.4 Числовi характеристики дискретного статистичного матерiалу.

1. **Вибiркова середня величина** *x*¯*B*. Величину, яка визначається за формулою:

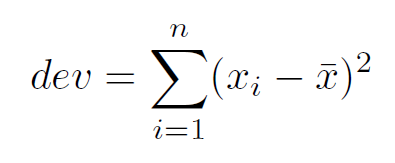


називають вибiрковою середньою величиною дискретного статистичного розподiлу вибiрки.

1. **Мода** (Мо\*). Модою дискретного статистичного розподiлу вибiрки називають варiанту, що має найбiльшу частоту появи.

Мод може бути кiлька. Коли дискретний i статистичний розподiл має одну моду, то вiн називається одномодальним, коли має двi моди - двомодальним i т.д.

1. **Медiана** (Ме\*). Медiаною дискретного статистичного розподiлу вибiрки називають варiанту, яка подiляє варiацiйний ряд на двi частини, рiвнi за кiлькiстю варiант;
2. **Девiацiя** - сума квадратiв вiдхилень елементiв статистичного матерiалу вiд середнього арифметичного.



1. **Варiансою** *s*2 називається девiацiя подiлена на обсяг статистичного матерiалу без одного.

A picture containing text, clock

Description automatically generated

1. **Стандартом** називається арифметийний корiнь з варiанси i позначається

A picture containing shape

Description automatically generated

1. **Дисперсiя**. Для вимiрювання розсiювання варiантв вибiрки вiдносно *x*¯*B* вибирається дисперсiя. Дисперсiя вибiрки - це середнє армифметичне квадратiв вiдхилень варiант вiдносно *x*¯*B* , яке обчислюється за формулою:

Text, letter

Description automatically generated

1. **Середнє квадратичне вiдхилення вибiрки** *σB*. При обчисленнi *DB* вiдхилення пiдноситься до квадрата, а отже змiнюється одиниця вимiру ознаки *X*, тому на основi дисперсiї вводиться середнє квадратичне вiдхилення.

A picture containing icon

Description automatically generated

яке вимiрює розсiювання варiант вибiрки вiдносно *x*¯*B*, але в тих самих одиницях, в яких вимiрюється ознака *X*;

1. **Розмах** (R), Для грубого оцiнювання розсiювання варiант вiдносно *x*¯*B* застосовується величина, яка дорiвнює рiзницi мiж найбiльшою *xmax* i найменшою *xmin* варiантами варiацiйного ряду. Ця величина називається розмахом

Text

Description automatically generated

1. **Квантилем порядку** *α*, якщо вiн iснує називається цей елемент статистичного матерiалу (вiдповiдного варiацiйного ряду), до якого включно маємо *α*% елементiв статистичного матерiалу (вiдповiдного варiацiйного ряду).

При *α < β* рiзницю мiж квантилем порядку *β* i квантилем порядку *α*

називають iнтерквантильною широтою порядку *β − α*

1. **Моментом** порядку *k* вiдносно сталої *a* називається вираз

A picture containing text, watch, clock

Description automatically generated

1. **Асиметрiєю** (*γ*1) або скошенiстю статистичного матерiалу називається вiдношення третього центрального моменту до другого центрального моменту в степенi пiвтора

A picture containing text, clock

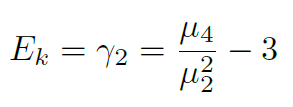
Description automatically generated

При *γ*1 *>* 0 бiльшiсть елементiв вибiрки зосереджено в лiвiй половині iнтервалу (статистичний матерiал скошений вправо).

При *γ*1 *<* 0 бiльшiсть елементiв вибiрки зосереджено в правiй половині iнтервалу (статистичний матерiал скошений влiво).

При *γ*1 = 0 статистичний матерiал розташований симетрично вiдносно середини iнтервалу.

1. **Ексцесом** (*γ*2) (крутiстю, сплющенiстю) статистичного матерiалу називається вiдношення четвертого центрального моменту до другого центрального моменту в квадратi мiнус три



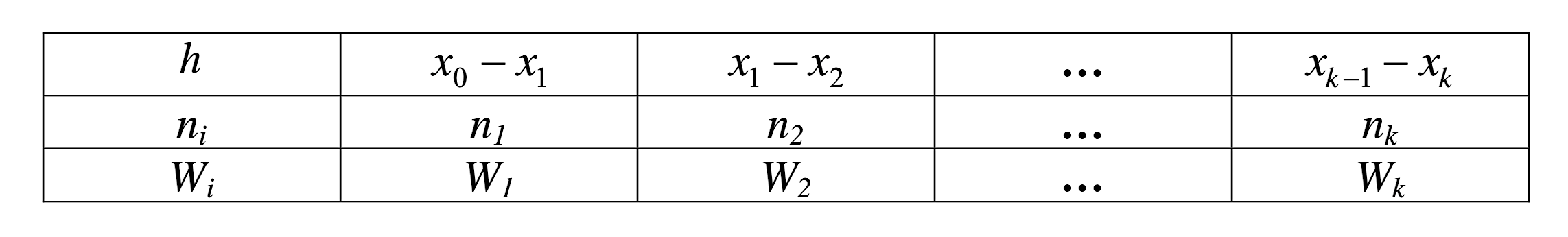
Вiн виражає ступiнь концентрацiї елеменiв вибiрки в околi її середнього:

Якщо *γ*2 *>* 0, то статистичний матерiал високовершинний; якщо *γ*2 *<* 0, то статистичний матерiал низьковершинний; якщо *γ*2 = 0, то статистичний матерiал нормально вершинний.

## Iнтервально статистичний розподiл вибiрки та його числовi характеристики.

Перелiк часткових iнтервалiв i вiдповiдних їм частот, або вiдносних частот називають iнтервальним статистичним розподiлом вибiрки.

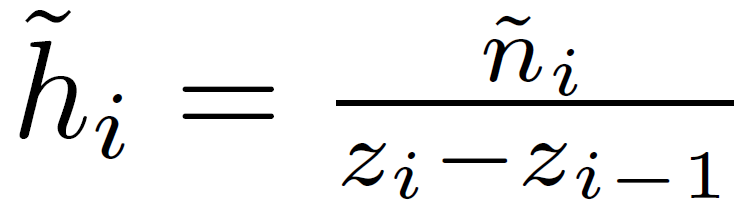
У табличнiй формi цей розподiл має такий вигляд:



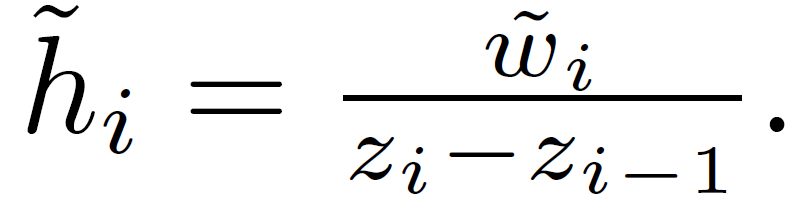
Тут *h* = *xi - xi*−1 є довжиною часткового iнтервалу. Як правило цей iнтервал береться однаковим.

* 1. Гістограма частот

Iнтервальний статистичний розподiл вибiрки можна подати графiчно у виглядi гiстограми частот або вiдносних частот, а також, як i для дискретного статистичного розподiлу, емпiричною функцiєю *F* ∗(*x*) (комулятою).

Гiстограмою частот називається схiдчаста фiгура, яка складена з прямокутникiв, основами яких є частиннi iнтервали (*xi*−1; *zi*]*, i* = 1*,* 2*..., m*, а їх висоти 

Площа кожного такого прямокутника дорiвнює *ni*.

Гiстограмою вiдносних частот називається схiдчаста фiгура, яка складена з прямокутникiв, основами яких є частиннi iнтервали (*zi*−1; *zi*], а їх висоти 

* 1. Емпiрична функцiя розподiлу.

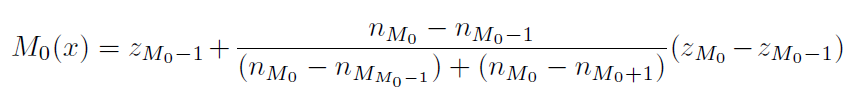
Iнтервально статистичний розподiл вибiрки також характеризується своєю емпiричною функцiєю розподiлу, але, на вiдмiну вiд дискретного випадку, вона геометрично зображається ламаною лiнiєю, яка з’єднує послiдовно точки (*zi, ωi*), де *ωi* = *w*1 + *w*2 + *...* + *wi, ω*0 = 0

* 1. Числовi характеристики iнтервального статистичного матерiалу.

1. Мода

У випадку iнтервального статистичного розподiлу визначають модальний iнтервал, тобто iнтервал [*zM*0 −1*, zM*0 ], якому вiдповiдає найбiльша частота *nM*0 .

Тодi моду обчислюємо у виглядi



1. Медіана

Щоб знайти медiану iнтервального статистичного розподiлу вибiрки, потрiбно спочатку видiлити медiанний iнтервал, тобто той частинний iнтервал [*xM*−1*, zM* ] злiва i справа вiд якого розмiщенi не бiльше половини варiант спостережень.

Нехай *nM* - вiдповiдна йому частота, а *mM*−1 накопичена частота попереднього iнтервалу. Тодi медiана

Text, letter

Description automatically generated

Програмна реалізація

Я реалізував програму використовуючи мову програмування Python в середовищі JupyterNotebook. З бібліотек я використав лише MatplotLib для зручної побудови графіків.

З самого початку генерується рандомна вибірка, можна задати її діапазон і кількість елементів. Далі обчислюються масиви значень без повторів (convert\_to\_set), відповідних ймовірностей (find\_probabilities) та відповідних кількостей (number\_of\_times\_occurs), скільки вони зустрічаються у вибірці. Саме через них працює більшістьфункцій.

Допоміжні функції: sort\_list, is\_in\_list, convert\_to\_set, number\_of\_times\_occurs, print\_frequency\_table, print\_empirical\_function, empirical\_function\_intervals, find\_probabilities, show\_plot, show\_plot\_intervals, sum\_, find\_max, find\_min

3.1 Обчислення числових характеристик дискретного розподілу

Як було написано вище, для дискретного розподілу я створив масиви для полегшення коду. Перший зберігає *xi*, а другий *ni.*

Функції для обчислення:

get\_median\_discrete, get\_moda\_discrete, get\_average, get\_scope, get\_deviation, get\_variance, get\_standart, get\_variation, get\_selective\_dispersion, get\_selective\_average\_quadratic\_deviation, get\_m\_k, get\_asymetry, get\_excess, get\_quantiles.

3.2 Обчислення числових характеристик інтервального розподілу

На перший погляд в інтервальному розподілі всі функції з дискретного мали б працювати так само, але це не так. В цьому розподілі є свої особливості, тому довелось трішки підправити кожну функцію.

Я написав особливу допоміжну функцію find\_intervals, яка розбиває варіаційний ряд на інтервали. Ці інтервали потім використовують інші функції. Також цікавою вийшла функція find\_right\_interval, яка повертає індекс інтервала потрібного для обчислення медіани і функція most\_modnyi для визначення масиву індексів найбільш модних інтервалів.

Функції: find\_intervals, is\_in\_tuple, find\_counts\_of\_elements, get\_midpoint, find\_right\_interval, find\_wi, index\_of\_max, return\_indexes\_of\_max, most\_modnyi, find\_moda\_intervals, get\_midpoint, average\_intervals, deviation\_intervals, variance\_intervals, standart\_intervals, m\_k\_intervals, selective\_dispersion\_intervals, show\_histogram\_intervals, selective\_average\_quadratic\_deviation\_intervals, asymetry\_intervals, excess\_intervals, empirical\_function\_intervals

Слово «intervals» в кінці назви кожної функції було допомогою для мене не збитись з пантелику при написанні коду. Таким чином я знав, що якщо функція, закінчується цим словом, то вона є для інтервального розподілу, а якщо ні, то для дискретного.

Отримані результати

Розмір вибірки: 50

Діапазон значень: 20

Згенерована вибірка: [17, 6, 14, 19, 14, 20, 11, 7, 10, 15, 9, 13, 14, 11, 19, 17, 6, 6, 5, 9, 19, 17, 15, 10, 16, 19, 9, 10, 14, 19, 6, 9, 9, 8, 14, 5, 17, 5, 12, 17, 12, 9, 15, 13, 14, 16, 6, 6, 18, 11]

Варіаційний ряд: [5, 5, 5, 6, 6, 6, 6, 6, 6, 7, 8, 9, 9, 9, 9, 9, 9, 10, 10, 10, 11, 11, 11, 12, 12, 13, 13, 14, 14, 14, 14, 14, 14, 15, 15, 15, 16, 16, 17, 17, 17, 17, 17, 18, 19, 19, 19, 19, 19, 20]

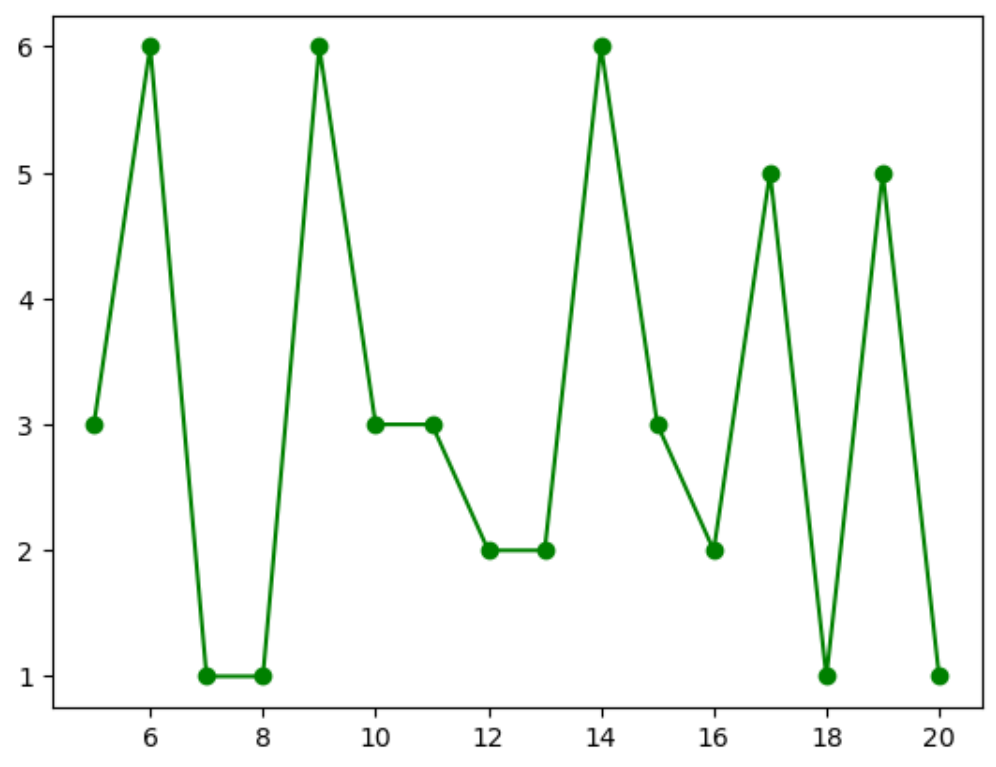
A picture containing table

Description automatically generated

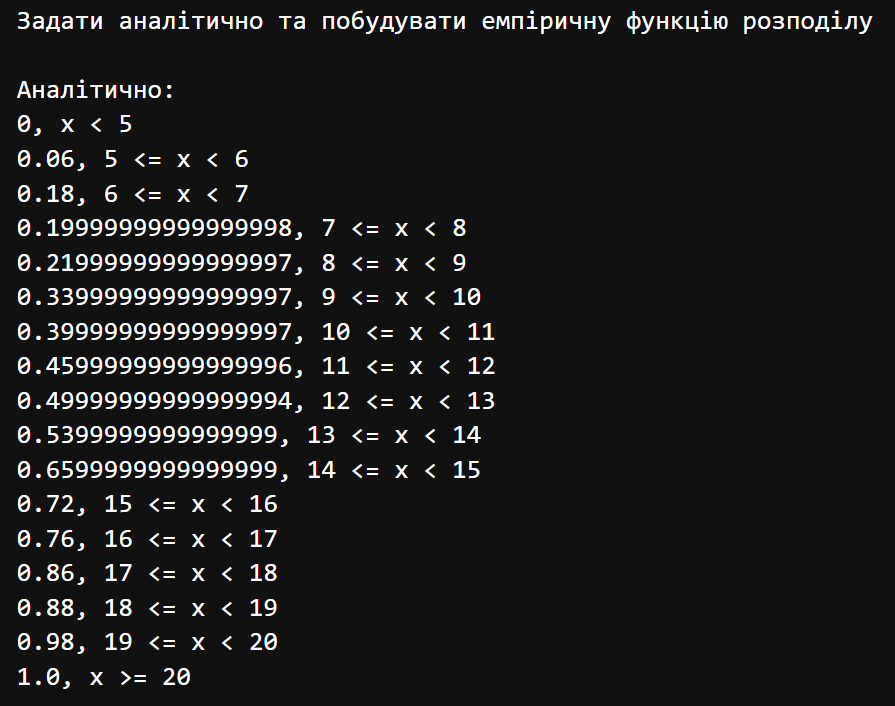
Chart, bar chart, histogram

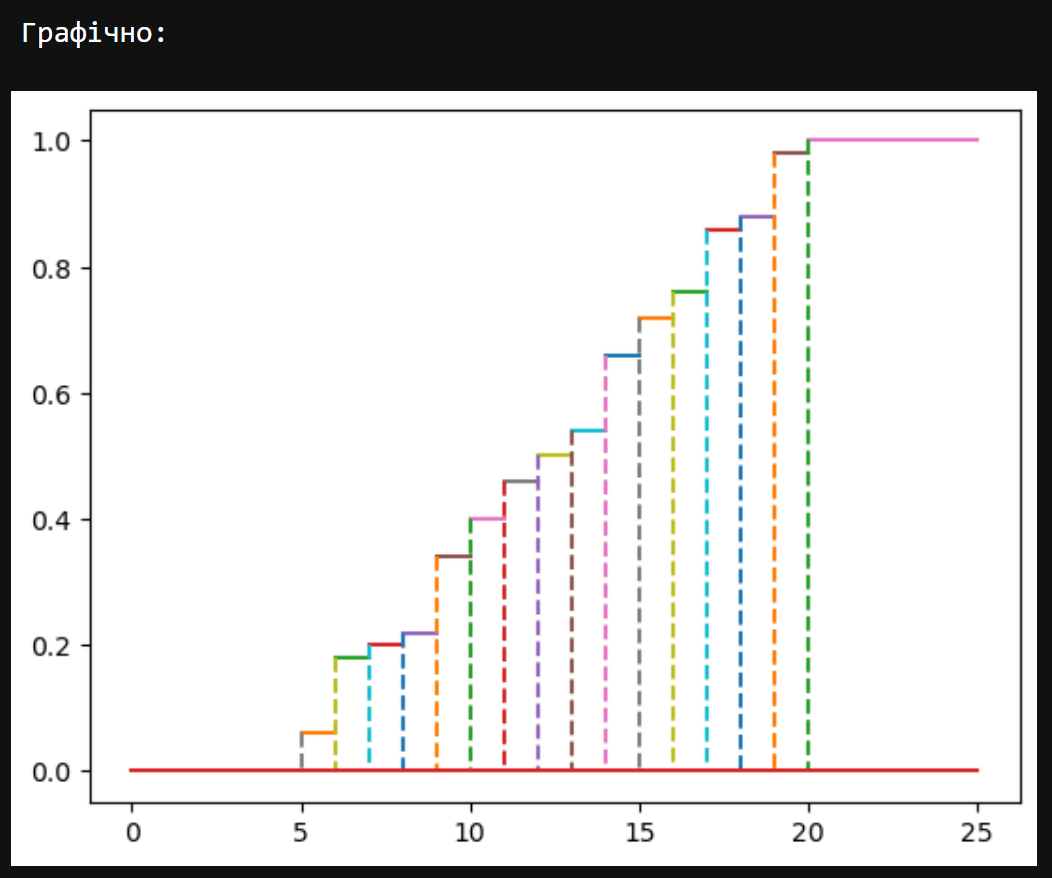
Description automatically generated

Полігон частот



Емпірична функція розподілу





Text

Description automatically generated

Асиметрія = -0,0247274 => статистичний матеріал скошений вліво.

Ексцес = -1,21444598 => статистичний матеріал низьковершинний.

Chart, line chart

Description automatically generated

Chart, histogram

Description automatically generated

A picture containing text

Description automatically generated

Text

Description automatically generated

Aсиметрія = -0.03930866 => статистичний матеріал скошений вліво.

Eксцес = -1.30556690 => статистичний матеріал низьковершинний.

Приклад великої вибірки (для демонстрації квантилів)

Розмір вибірки: 1000

Діапазон значень: [5; 25]

Згенерована вибірка: [21, 23, 10, 22, 6, 14, 8, 5, 12, 12, …, 10, 10, 12, 11, 11, 5, 17, 16]

Варіаційний ряд: [5, 5, 5, 5, 5, 5, 5, 5, 5, …, 25, 25, 25, 25, 25]

A screenshot of a computer

Description automatically generated with low confidence

Chart, bar chart, histogram

Description automatically generated

Chart, line chart

Description automatically generated

Text

Description automatically generated

Chart, histogram

Description automatically generated

Text

Description automatically generated

Aсиметрія = 0.049265 => статистичний матеріал скошений вправо.

Eксцес = -1.2702905 => статистичний матеріал низьковершинний.

Background pattern

Description automatically generated

Text, logo

Description automatically generated

Text

Description automatically generated

Aсиметрія = 0.050095 => статистичний матеріал скошений вправо.

Eксцес = -1.2894623 => статистичний матеріал низьковершинний.

A picture containing calendar

Description automatically generated

Chart, line chart

Description automatically generated

Chart, histogram

Description automatically generated

Висновок

Під час цієї індивідуальної роботи я навчився будувати варіаційний ряд, будувати частотну таблицю та полігон частот, будувати графіки статистичного матеріалу, будувати емпіричну функцію розподілу та обчислювати всі числові характеристики для інтервального та дискретного розподілів (медiану, моду, середнє вибiркове, девiацiю, варiансу, дисперсiю, стандарт, варiацiю, розмах, квантилi, центральні моменти, асиметрiю та ексцес).