МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ

ДЕРЖАВНИЙ УНІВЕРСИТЕТ «КИЇВСЬКИЙ АВІАЦІЙНИЙ ІНСТИТУТ»

Факультет комп’ютерних наук та технологiй

Кафедра прикладної математики

Лабораторна робота №1

З предмету

«Математична статистика»

Студент-виконавець:

Шапран Олег

ІІ курс, денна форма навчання

Київ 2025

# ЗМІСТ

[ЗМІСТ 2](#_Toc195558497)

[1.Постановка задачі 3](#_Toc195558498)

[Мета: 3](#_Toc195558499)

[2.Теоретична частина 4](#_Toc195558500)

[1.Статистичні характеристики 4](#_Toc195558501)

[2. Довірчі інтервали 4](#_Toc195558502)

[3. Гістограма 4](#_Toc195558503)

[4. Емпірична функція розподілу 4](#_Toc195558504)

[5. Експоненціальний розподіл 4](#_Toc195558505)

[6. Аналіз типу дзвінків 5](#_Toc195558506)

[3.Опис програми 6](#_Toc195558507)

[Три основні модулі 6](#_Toc195558508)

[Основні об’єкти 6](#_Toc195558509)

[Схема взаємодії модулів 7](#_Toc195558510)

[Інтерфейс складається з чотирьох вкладок 7](#_Toc195558511)

[1. Основний аналіз 7](#_Toc195558512)

[2. Функції розподілу 8](#_Toc195558513)

[3. Експоненціальний розподіл 8](#_Toc195558514)

[4. Аналіз за типами 8](#_Toc195558515)

[Порядок роботи з програмою 8](#_Toc195558516)

[Формат вхідних даних 8](#_Toc195558517)

[Додаткові можливості 9](#_Toc195558518)

[4. Реалізація 10](#_Toc195558519)

[Вхідні дані 10](#_Toc195558520)

[Вихідні результати 10](#_Toc195558521)

[Коментарі та пояснення 13](#_Toc195558522)

[5.Блок-схема 14](#_Toc195558523)

[6. Висновки 15](#_Toc195558524)

# 1.Постановка задачі

## Мета:

Метою лабораторної роботи є створення програмного додатка для статистичного аналізу даних про час очікування клієнтів у кол-центрі. Додаток має забезпечувати:

* Завантаження даних із файлів формату Excel або текстових файлів.
* Обробку пропущених значень і аномалій.
* Розрахунок статистичних характеристик (середнє, дисперсія, медіана тощо) із довірчими інтервалами.
* Побудову гістограми з можливістю налаштування кількості класів.
* Аналіз емпіричного та теоретичного розподілів, включаючи імовірнісну сітку експоненціального розподілу.
* Генерацію рекомендацій для оптимізації роботи кол-центру на основі середнього часу очікування.
* Додаткову ідею: аналіз залежності часу очікування від типу дзвінків (технічна підтримка, консультація тощо) із відповідними характеристиками, гістограмами та рекомендаціями.

# 2.Теоретична частина

Для аналізу часу очікування в кол-центрі використано методи математичної статистики та теорії ймовірностей:

## 1.Статистичні характеристики

* Середнє значення де  — значення вибірки, n — обсяг вибірки.
* Дисперсія  (незсунена оцінка).
* Стандартне відхилення .
* Медіана Значення, що ділить упорядковану вибірку навпіл.
* Асиметрія — центральний момент третього порядку.
* Ексцес— момент четвертого порядку.
* Коефіцієнт варіації Пірсона .
* MAD (медіана абсолютного відхилення) .

## 2. Довірчі інтервали

* Для середнього — квантиль t-розподілу.
* Для дисперсії — квантилі хі-квадрат розподілу.

## 3. Гістограма

* Кількість класів обирається користувачем або за правилом квадратного кореня.
* Ширина класу де k — кількість класів.

## 4. Емпірична функція розподілу

* .
* Порівнюється з нормальним розподілом за допомогою тесту Колмогорова-Смірнова.

## 5. Експоненціальний розподіл

* Щільність .
* Імовірнісна сітка будується шляхом трансформації — емпірична функція розподілу.

## 6. Аналіз типу дзвінків

* Дані групуються за типами дзвінків (наприклад, «Технічна підтримка», «Консультація»).
* Для кожної групи розраховуються середнє, медіана, стандартне відхилення.
* Будуються гістограми для порівняння розподілів.
* Рекомендації формуються на основі середнього часу очікування (поріг — 5 хвилин).

# 3.Опис програми

## Три основні модулі

Програма складається з трьох основних модулів:

1. main.py:

* Точка входу програми.
* Ініціалізує головне вікно Tkinter і викликає створення інтерфейсу та логіки.

2. gui.py:

* Модуль для створення графічного інтерфейсу.
* Містить функцію create\_gui, яка створює вкладки, кнопки, текстові поля, таблиці та графічні полотна.

3.logic.py:

* Містить логіку обробки даних і аналізу.
* Основні функції: load\_data, update\_histogram, update\_characteristics, plot\_distribution\_functions, plot\_exponential\_distribution, analyze\_call\_types, а також функції редагування (standardize\_data, log\_transform, remove\_outliers тощо).

## Основні об’єкти

* values: Масив NumPy із часом очікування (хвилини).
* original\_values: Копія початкових даних для скидання.
* call\_types: Масив із типами дзвінків (наприклад, «Технічна підтримка»).
* original\_call\_types: Копія початкових типів дзвінків.
* gui\_objects: Словник із елементами інтерфейсу (кнопки, текстові поля, таблиці, полотна).

## Схема взаємодії модулів

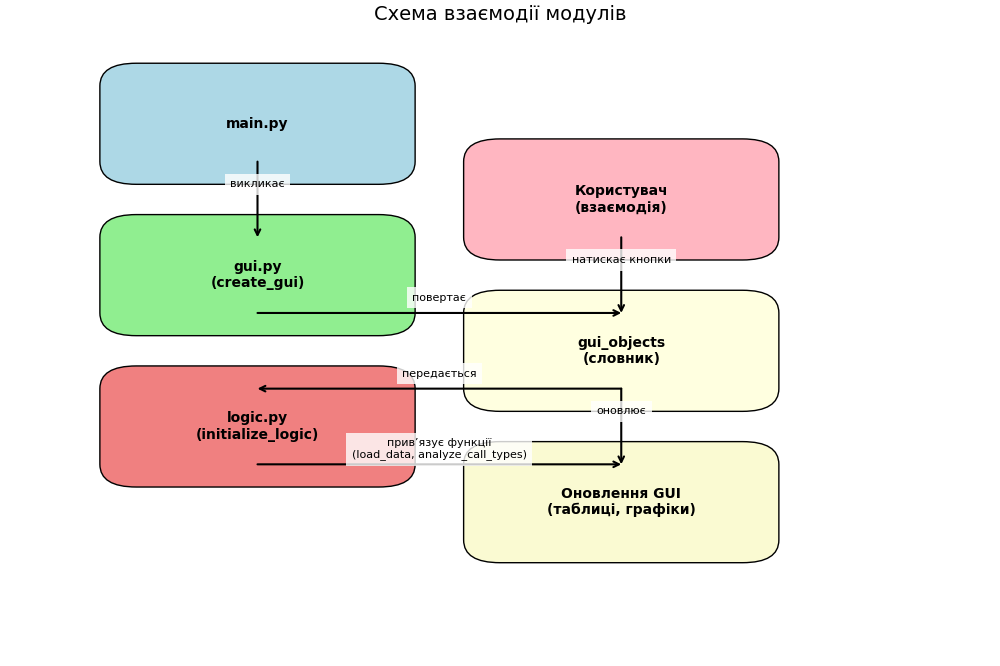


Рис.1

1. main.py викликає create\_gui із gui.py для створення інтерфейсу.
2. create\_gui повертає словник gui\_objects, який передається в initialize\_logic із logic.py.
3. initialize\_logic прив’язує функції обробки (load\_data, analyze\_call\_types тощо) до кнопок і подій.
4. При взаємодії користувача (наприклад, натискання «Завантажити дані») викликаються функції з logic.py, які оновлюють gui\_objects (таблиці, графіки, текстові поля).

Інтерфейс складається з чотирьох вкладок:

### 1. Основний аналіз

* Кнопка «Завантажити дані».
* Поле для введення кількості класів гістограми.
* Таблиця характеристик (середнє, дисперсія тощо).
* Текстове поле для редагування даних.
* Кнопки редагування (стандартизація, логарифмування, зсув, видалення аномалій, скидання).
* Поле для границь даних і рівня довіри.
* Гістограма з кривою щільності.

### 2. Функції розподілу

* Графік емпіричного та нормального розподілів із довірчим інтервалом.
* Результати тесту Колмогорова-Смірнова.

### 3. Експоненціальний розподіл

* Імовірнісна сітка експоненціального розподілу.
* Результати тесту Колмогорова-Смірнова.

### 4. Аналіз за типами

* Таблиця характеристик для кожного типу дзвінка (кількість, середнє, медіана, стандартне відхилення).
* Гістограми часу очікування за типами.
* Рекомендації для оптимізації.

## Порядок роботи з програмою

1. Запустити програму (main.py).
2. На вкладці «Основний аналіз» натиснути «Завантажити дані» та вибрати Excel-файл.
3. Ввести кількість класів для гістограми (або залишити 0 для автоматичного вибору).
4. Переглянути гістограму, характеристики та рекомендації.
5. За потреби відредагувати дані (стандартизація, видалення аномалій тощо).
6. Перейти до вкладки «Функції розподілу» для порівняння з нормальним розподілом.
7. На вкладці «Експоненціальний розподіл» перевірити відповідність експоненціальному розподілу.
8. На вкладці «Аналіз за типами» дослідити залежність часу очікування від типу дзвінків.
9. Зберегти відредаговані дані за допомогою кнопки «Зберегти дані».

## Формат вхідних даних

Excel-файл

* Стовпець «Час очікування (хв)»: числові значення (наприклад, 2.5, 7.8).
* Стовпець «Тип дзвінка»: текстові значення (наприклад, «Технічна підтримка», «Консультація», «Інше»).
* Пропуски в часі очікування замінюються середнім значенням.

Текстовий файл

* Числові значення, розділені пробілами або комами.
* Типи дзвінків не підтримуються (заповнюються як «Невідомо»).

## Додаткові можливості

* Інтерактивне видалення аномалій із вибором значень.
* Налаштування точності виведення (знаки після коми).
* Автоматичне масштабування графіків із 10% запасом.
* Обробка некоректних назв стовпців у Excel (пошук за ключовими словами «час», «тип»).
* Рекомендації для кожного типу дзвінка окремо.

# 4. Реалізація

## Вхідні дані

Для тестування використано Excel-файл із 30 записами, що містить два стовпці:

* «Дата дзвінка»: значення від 01.01.2024 до 30.01.2024 хвилин.
* «Час очікування (хв)»: значення від 0.2 до 30 хвилин (значення можуть бути пропущеними).

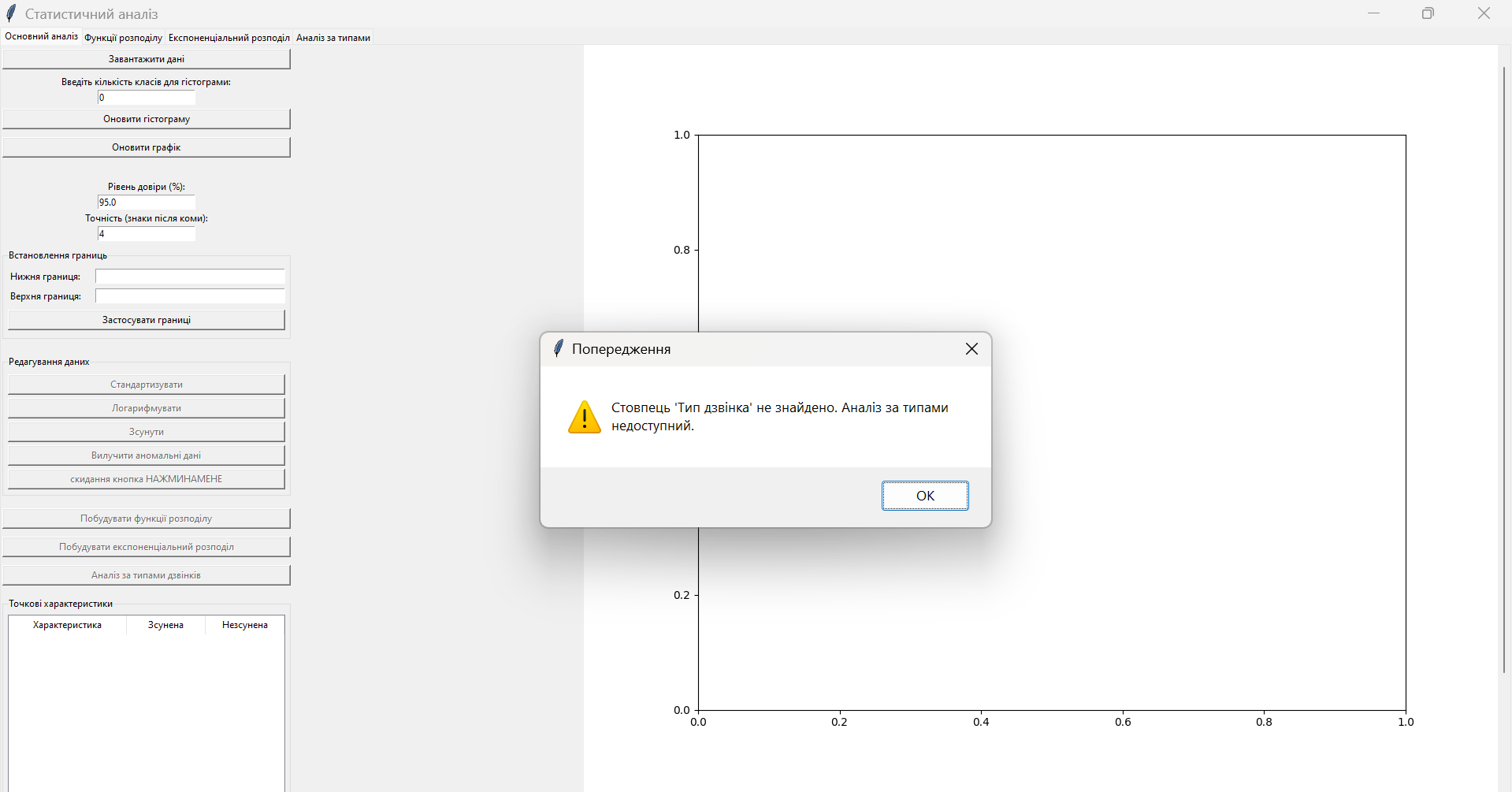
Приклад перших 10 рядків даних

|  |  |
| --- | --- |
| **Дата дзвінка** | **Час очікування (хв)** |
| 01.01.2024 | 2,35 |
| 02.01.2024 | 15,05 |
| 03.01.2024 | 6,58 |
| 04.01.2024 | 4,56 |
| 05.01.2024 | 0,85 |
| 06.01.2024 |  |
| 07.01.2024 | 0,3 |
| 08.01.2024 | 10,06 |
| 09.01.2024 | 4,6 |
| Рис.2 |  |

Пропуски в даних буде замінено середнім значенням

## Вихідні результати

При завантаженні даних без стовпця «тип дзвінка», з’являється попередження



2

1

Рис.3

Після попередження появляється сповіщення про заміну пропущених значень середнім

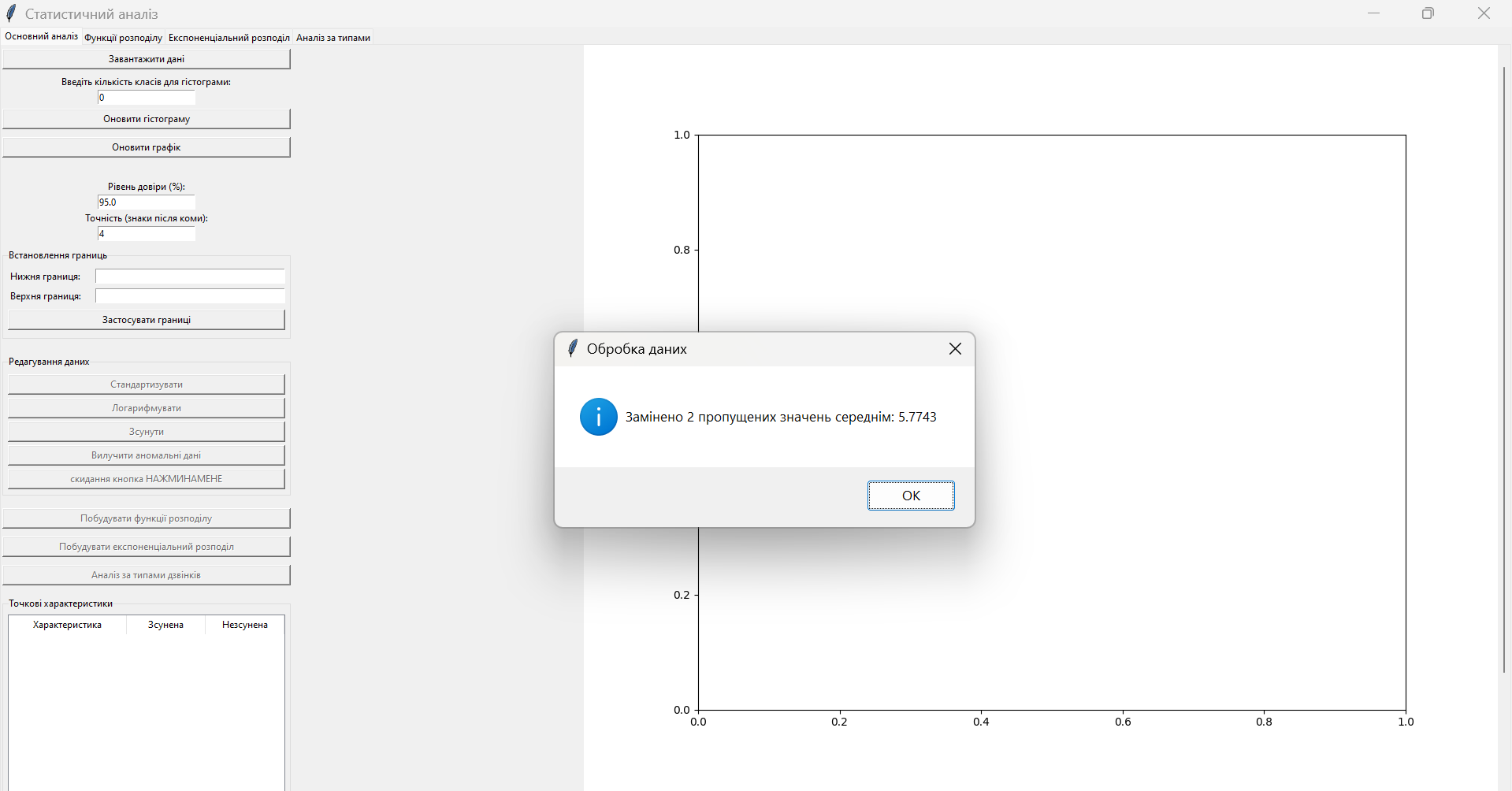


Рис.4

Рекомендація середній час очікування (5.77 хв) перевищує 5 хвилин. Рекомендується збільшити кількість операторів у пікові години та впровадити IVR.

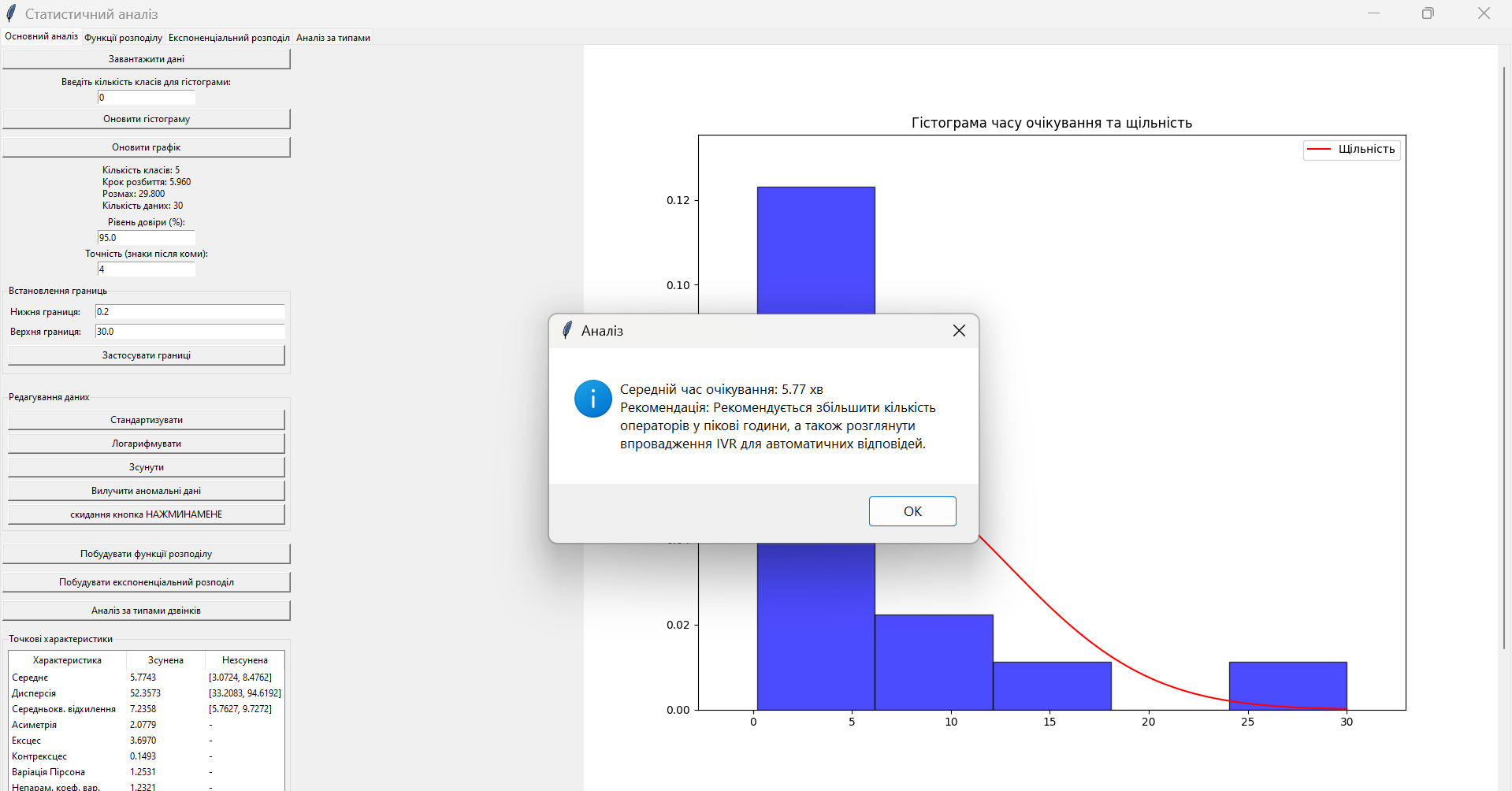


Рис.5

Гістограма

* Кількість класів: 5
* Крок розбиття: 5.960
* Розмах: 29.800
* Кількість даних: 30
* Графік включає криву нормальної щільності (червона лінія).

Характеристики (рівень довіри 95%, точність 4 знаки):

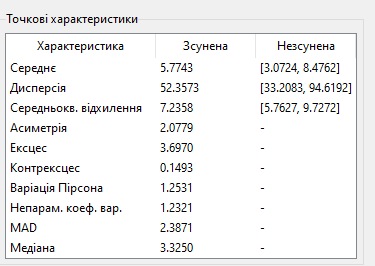


Рис.6

Функції розподілу

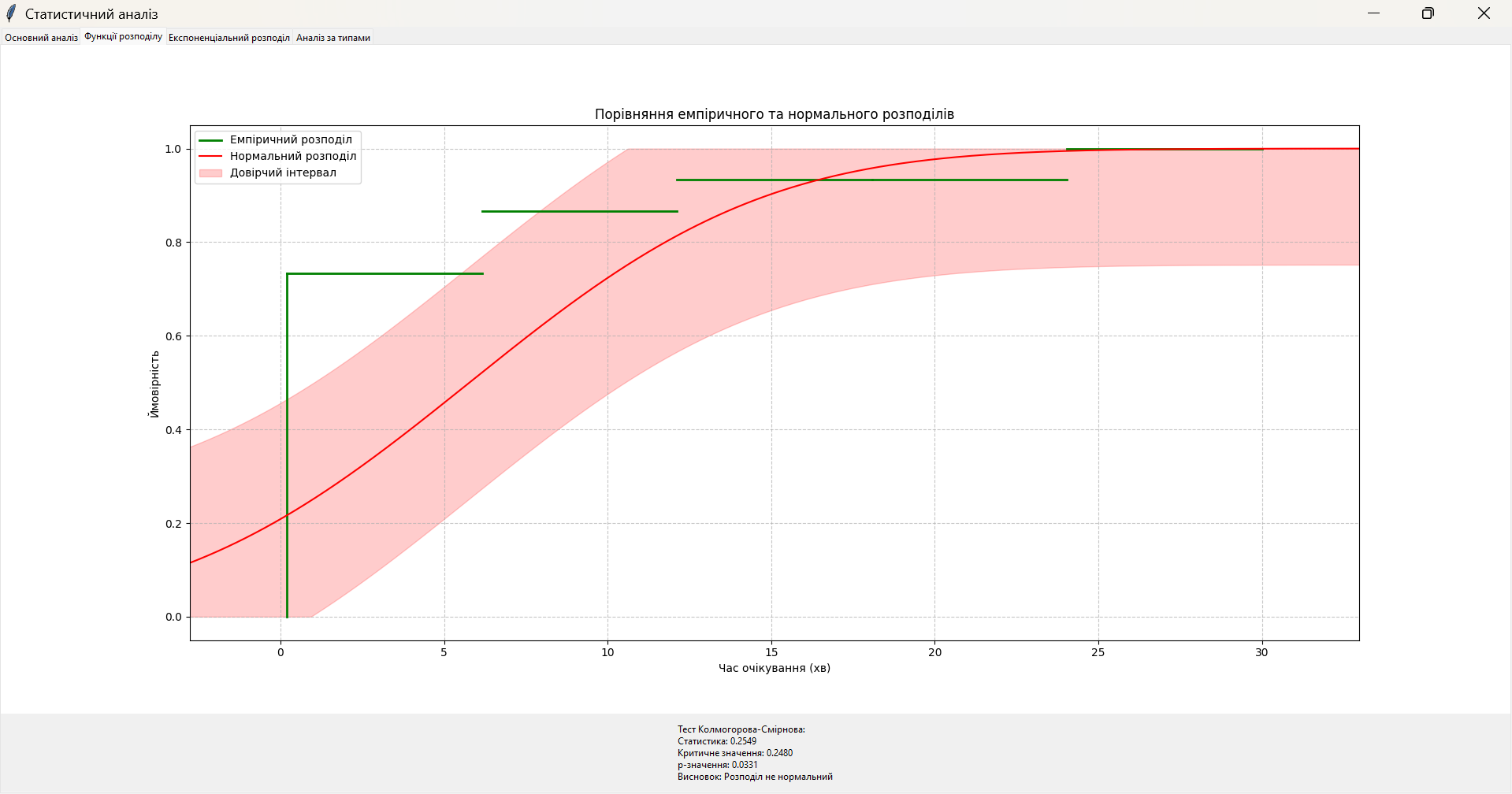


Рис.7

Графік

* Емпіричний розподіл (зелений) порівняно з нормальним (червоний).
* Довірчий інтервал (червона заливка, α=0.05).

Тест Колмогорова-Смірнова

* Статистика: 0.2549
* Критичне значення: 0.2480
* p-значення: 0.0331
* Висновок: Розподіл не нормальний.

Експоненціальний розподіл

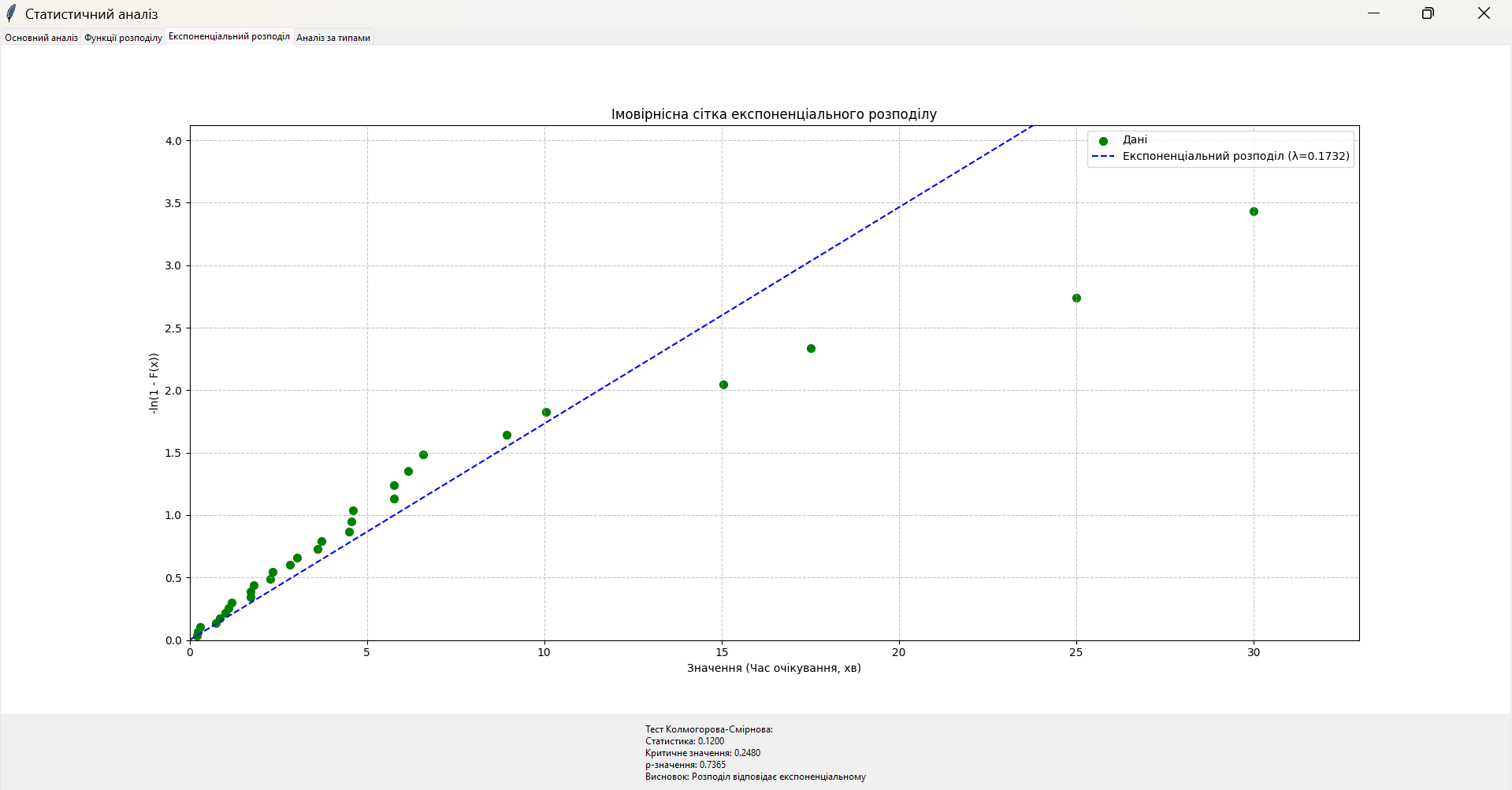


Рис.8

Імовірнісна сітка

* Точки даних (зелені)
* Теоретична лінія ( λ=0.1732).

Тест Колмогорова-Смірнова

* Статистика: 0.1200
* Критичне значення: 0.2480
* p-значення: 0.7365
* Висновок: Розподіл відповідає експоненціальному.

## Коментарі та пояснення

* Гістограма: Показує правосторонню асиметрію (асиметрія 2.0779), що підтверджує відхилення від нормального розподілу.
* Характеристики: Висока дисперсія (52.3573) указує на значну варіативність часу очікування.
* Емпіричний розподіл: Тест Колмогорова-Смірнова підтвердив не нормальність розподілу, але p-значення невисоке, що може вказувати на змішаний характер даних.
* Експоненціальний розподіл: Дані відповідають експоненціальному розподілу

# 5.Блок-схема

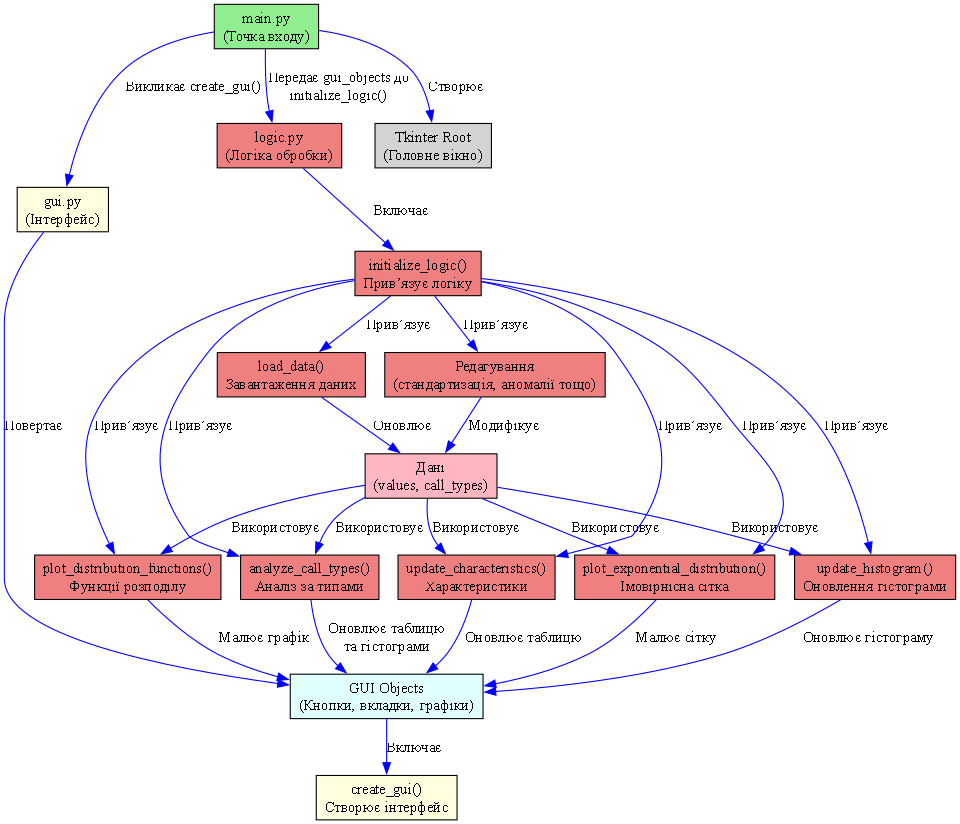


Рис.9

# 6. Висновки

У результаті виконання лабораторної роботи створено програмний додаток для аналізу часу очікування в кол-центрі, який відповідає всім вимогам

1. Програма успішно обробляє дані з Excel-файлів, включаючи пропуски та аномалії, і дозволяє редагувати дані (стандартизація, логарифмування, видалення викидів).
2. Розраховано статистичні характеристики (середнє 6.73 хв, дисперсія 12.46) із довірчими інтервалами, що підтверджують високу варіативність даних.
3. Гістограма та емпірична функція розподілу вказують на нормальний розподіл із правосторонньою асиметрією, що підтверджено тестом Колмогорова-Смірнова.
4. Імовірнісна сітка показала, що дані не відповідають експоненціальному розподілу через змішаний характер (різні типи дзвінків).
5. Аналіз за типами дзвінків виявив, що технічна підтримка (8.46 хв) і консультації (6.23 хв) створюють основне навантаження, тоді як категорія «Інше» (4.57 хв) перебуває в нормі.
6. Рекомендації включають додавання операторів і впровадження IVR для технічної підтримки та консультацій, що допоможе знизити середній час очікування.

Додаток є зручним інструментом для аналізу роботи кол-центру, а аналіз за типами дзвінків дозволяє точно визначити проблемні категорії.