深 圳 大 学 实 验 报 告

课程名称：**Probability & Statistics**

实验项目名称：**Experiment 4**

学院：**College of Electronic and Information Engineering**

专业：**Electronic Information Engineering**

指导教师：**Changsheng CHEN**

报告人：**Zuyi LIAO** 学号：**2022110131**

班级：**Wenhua Honors Class**

实验时间：**2023.12.29**

实验报告提交时间：**2023.12.29**

教务处制

Aim of Experiment:

- Familiar with the central limit theorem.
- Understand the implementation of the central limit theorem in python.
- Know how to visualize data in different distributions.

Experiment Content:

1. Demonstration of CLT in action using simulations in Python

Exercise 1: Exponentially distributed population

For a function $f(x) = \begin{cases} \frac{1}{\theta}e^{-\frac{x}{\theta}} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$ assuming that θ=4 , please calculate the mean and the standard deviation of the population and check whether the CLT holds.

Exercise 2: Binomially distributed population

Given function $P(x) = \begin{cases} \binom{k}{x}(p)^x(1-p)^{1-x} & \text{if } x = 0, 1, 2, \ldots, k \\ 0 & \text{otherwise} \end{cases}$ we follow a similar approach

and plot the sampling distribution obtained with a large sample size (n=500) for a Binomially distributed variable with parameters k=30 and p=0.9.

2. An Application of CLT in Investing/Trading

Exercise:

In finance models (like the Black-Scholes option pricing model), we usually assume that the price series is log-normally distributed, and so returns are normally distributed. Please try to do a visual analysis of the distribution of daily returns.

3. More visualization exercises

Exercise:

In this assignment we will use the hourly weather dataset and look at the temperature recording for the city of Detroit. We will use continuous distributions we have studied to approximate the observed distributions .

Experiment Process：

1. **Demonstration of the Central Limit Theorem (CLT) Using Python Simulations**

This section includes experiments demonstrating that for any population distribution, the sampling distribution (of sample means) will tend to be normally distributed when the sample size is sufficiently large.

- **Exercise 1: Exponentially Distributed Population**

The objective is to calculate the theoretical mean and standard deviation for an exponentially distributed population (both of which are theta) and simulate this population to obtain the mean and standard deviation. The effect of varying sample sizes on the distribution will also be analyzed.

- **Exercise 2: Binomially Distributed Population**

This exercise is analogous to Exercise 1 but with a binomially distributed population. The mean is defined as n*p and the standard deviation as sqrt(n*p*(1-p)). The same analysis approach will be applied.

2. **Application of CLT in Investing/Trading**

- **Exercise:**

Utilizing the seaborn package, this exercise involves visualizing the returns of an investment or trading strategy and their distribution. This includes plotting a histogram with an automatically determined bin size.

3. **Additional Visualization Exercises**

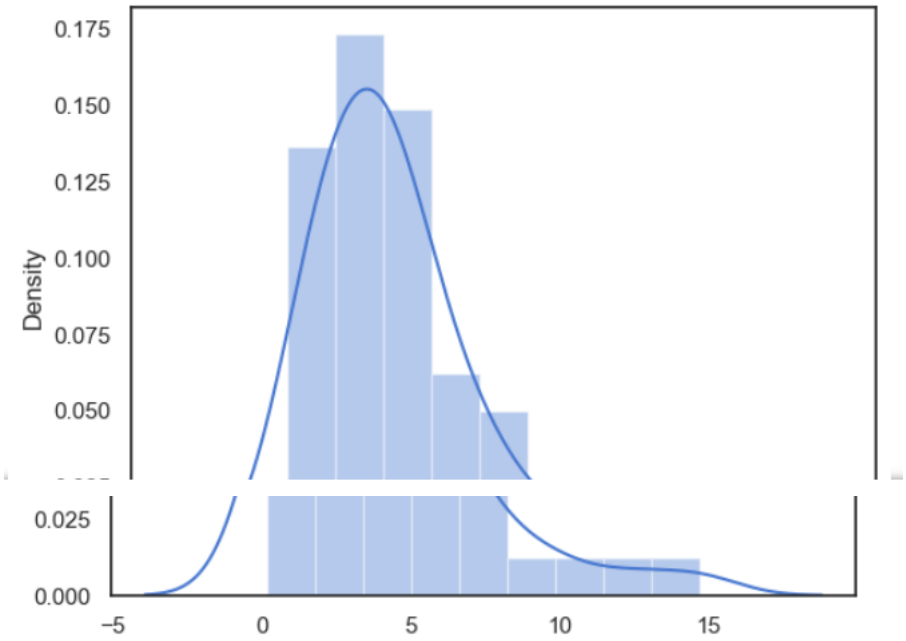
These exercises involve extensive research and troubleshooting, including addressing various errors encountered during experimentation. This process is both challenging and educational. I learn a lot.

Data Logging and Processing:

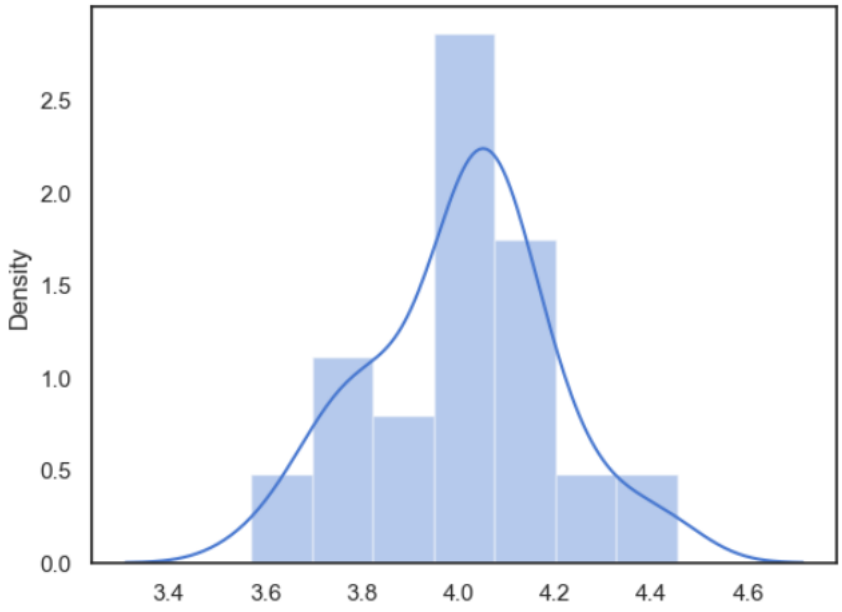
1. Demonstration of CLT in action using simulations in Python

Exercise 1:

Initially, I simulated a sample size of 50 and observed a noticeable deviation from the standard distribution. However, increasing the sample size to 500, the empirical values closely approximated the theoretical values. To better illustrate this, I created a figure that conveniently compares the means and standard deviations across different sample sizes.



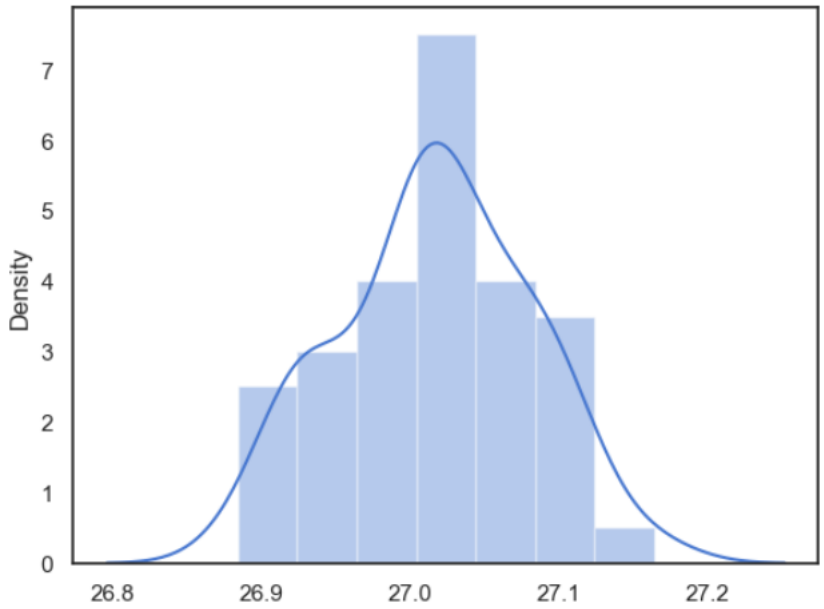
When we set n to 500, as can be seen from the figure, the empirical values are approximately equal to the theoretical values.



Exercise 2:

I drew 50 random samples, each of size 500, from a Binomial distribution with parameters

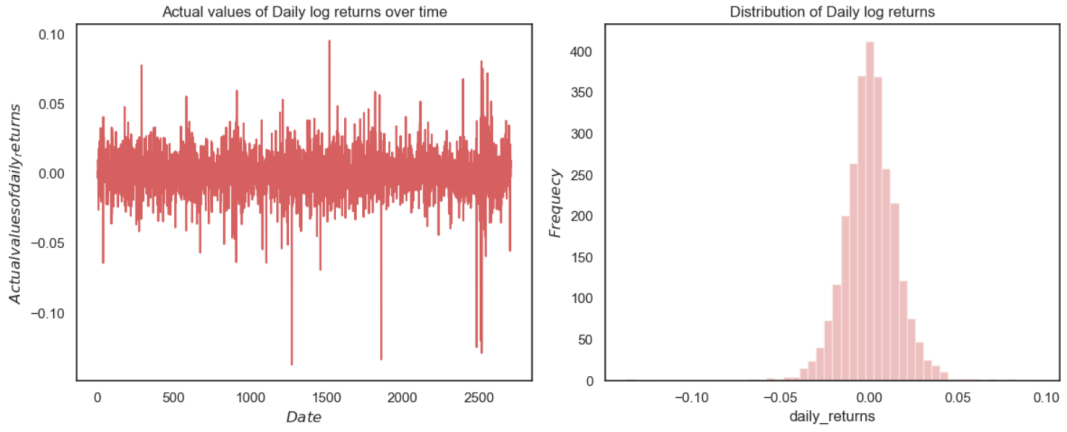
k=30 and p=0.9. The exercise involved calculating and comparing both the empirical and theoretical means and standard deviations, as depicted in the accompanying figure.



2. An Application of CLT in Investing/Trading

Exercise:

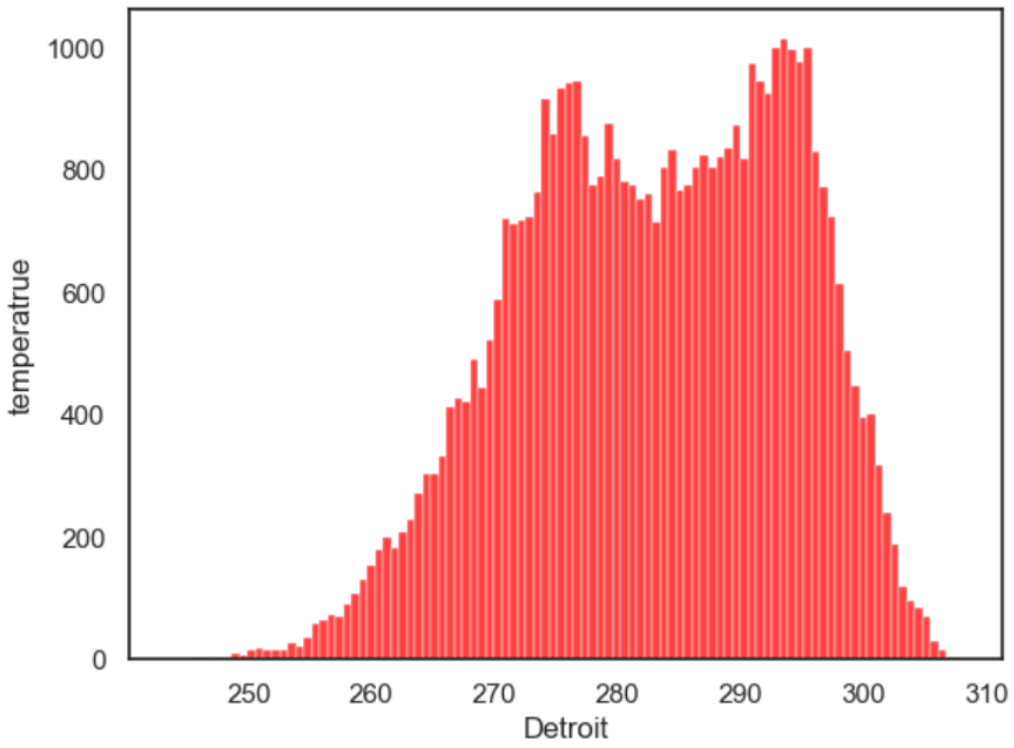
Analyzing ten years of daily returns for ITC, it's observed that they are centered around 0, with occasional spikes beyond five percent and drops below negative ten percent. This pattern, indicative of higher kurtosis or fat-tails, differs from a normal distribution and results in a leptokurtic distribution plot.



3. More visualization exercises

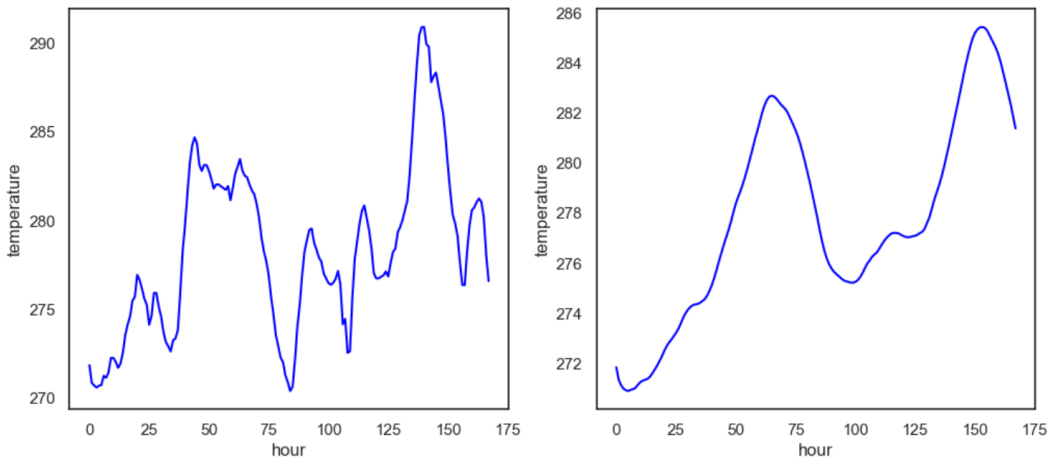
Exercise 1:

I plotted a histogram with 100 bins using all samples, including a slider to adjust the number of bins (n). Increasing n enhances the graph's accuracy.



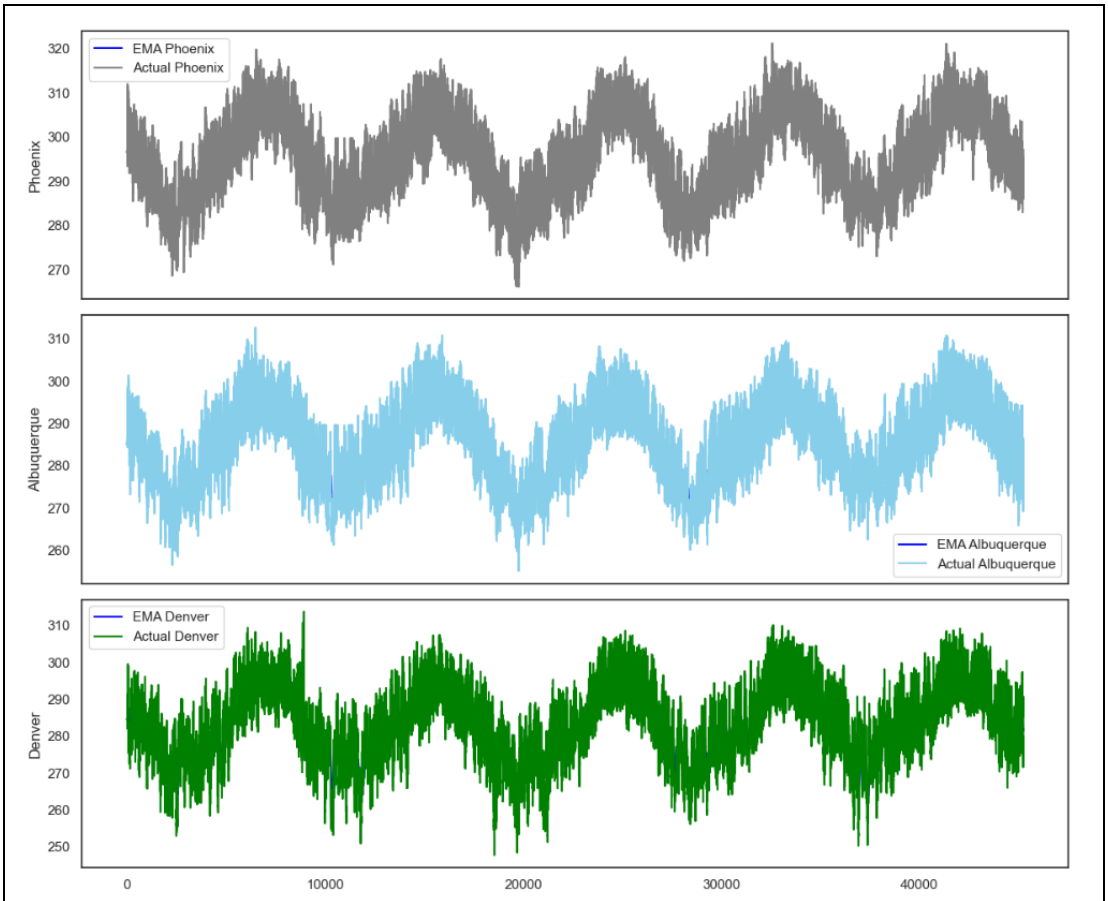
Exercise 2: 7-Day Moving Average Plot(#1)

The moving average method was used to predict the next cycle's data based on a certain cycle's data. The cycle length impacts the prediction accuracy; shorter cycles yield more real-time data responsiveness.



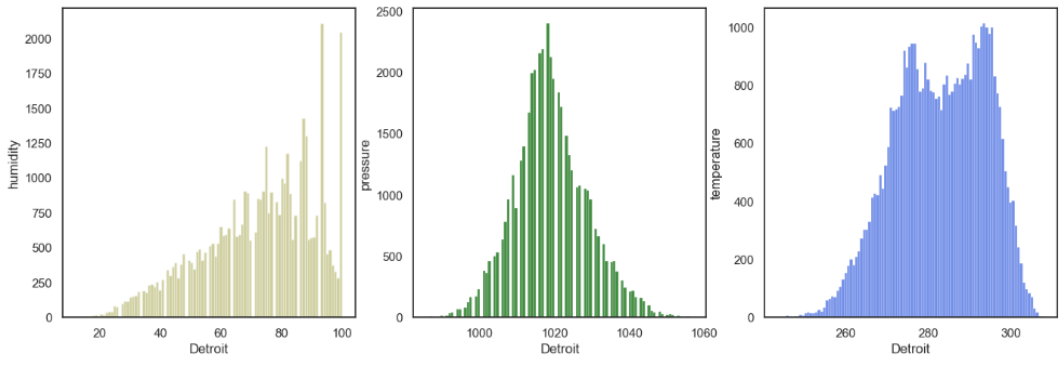
Exercise 3: Temperature Variation of Different Cities(#2)

This exercise involved plotting a graph with three subgraphs, sharing the same X-axis, to clearly illustrate each city's temperature trend. The graph shows a distinct periodicity in temperature changes across different cities.



Exercise 4: Visualization of Humidity, Pressure, and Temperature in Detroit(#3)

In this exercise, I plotted Detroit's humidity, pressure, and temperature trends in a single figure. The selected timeframe shows an upward trend in humidity, an initial increase followed by a decrease in pressure, and a more complex pattern in temperature changes. The figure provides detailed insights into these trends.



Experimental Results and Analysis:

- **Mastering Key Tools**:
  Deepened my understanding of Matplotlib, NumPy, and Seaborn, enhancing my skills in data analysis and visualization in Python.
- **Interactive Visualization**:
  Learned to create dynamic charts with sliders, improving the interactivity of data presentations.
- **Practical Application of Charts**:
  Through various chart exercises, I deepened my understanding of data visualization techniques and enhanced my ability to apply theory to practice.
- **Applying Moving Averages**:
  Applied moving average methods to real-world problems, enhancing my data processing and problem-solving skills.
- **Simulation and Statistical Understanding**:
  Strengthened my grasp of the Central Limit Theorem and statistical concepts through simulation experiments, learning to apply these principles in complex scenarios.
- **Data Analysis Skills**:
  Gained a preliminary mastery of data visualization and analysis, improving my ability to interpret and present data.

指导教师批阅意见：   成绩评定：   指导教师签字： 年　　月　　日
备注：

注：1、报告内的项目或内容设置，可根据实际情况加以调整和补充。

2、教师批改学生实验报告时间应在学生提交实验报告时间后 10 日内。