

Homework 1

We load data on monthly stock returns from 1926 to 2021 ([source](#)).

```
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

cd /content/drive/MyDrive/Intro to DS lecture

/content/drive/MyDrive/Intro to DS lecture

import numpy as np
import pandas as pd

stocks = pd.read_csv("stocks.csv")
ret = stocks["Mkt-RF"].values
```

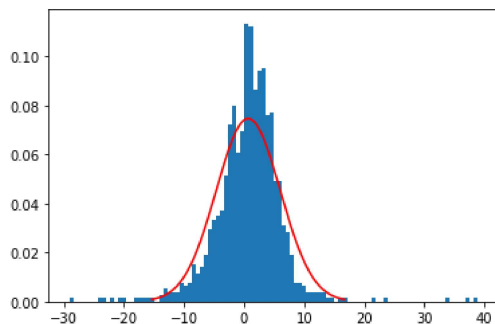
A *density histogram* is a histogram that is normalized so that the total area under the bars sums to unity. Argue that the stock return is not normally distributed as follows:

1. Plot a density histogram of the returns `ret` with 100 bins. [2 pts]
2. Over this histogram, plot the PDF of a normal distribution with mean equal to the sample mean of the returns and standard deviation equal to the sample standard deviation of the returns. [2 pts]

```
import matplotlib.pyplot as plt
import scipy.stats as stats

mu = np.mean(ret) # calculate mean to get mu
sigma = np.std(ret) # get sigma
ans = np.linspace(mu - 3*sigma, mu + 3*sigma, 1000)

plt.hist(ret, bins=100, density=True) #plot ret, set 100 bins
plt.plot(ans, stats.norm.pdf(ans, mu, sigma), color='red') # plot the PDF
plt.show()
```



3. Explain why stock returns are not normally distributed. [2 pts]

Answer: One of the simplest explanations is the unpredictability of the market. After all, the analysis of this stock is not a coin toss. Tossing a coin has a stable probability so it can make a perfect normal distribution curve. In the analysis of stocks, there will be some extreme values or large fluctuations. As we can see in the graph, extreme events occur more frequently in reality than the normal distribution predicts, so this result occurs. So we know that if we use the assumed normal distribution to analyze stocks, the result will be disastrous.

