Lab section 01

EXERCISE 1: BASIC DESCRIPTIVE STATISTICS

Download data

Google "*yahoo finance*", go to the web page. In the search bar, type the name of your favorite publically traded stock. Click on "*historical data"* to retrieve price data. On the next page, set the date range from January 1st, 2006 to January 1st, 2023, and click on *"monthly"*.

Yahoo does not always allow to download the data directly (for example I could do it for Amazon, but not for the SP500). If there is no download button shown on the webpage, you will need to either copy paste the table on a spreadsheet, or use a web scraping tool (I used the following Chrome Extension, Instant Data Scraper, which worked well).

Now re-do the same operation for the SP500, whose symbol on Yahoo Finance is "^GSPC". Consolidate the two tables in the same spreadsheet.

In the spreadsheet you created:

Open refers to the "opening price," i.e., to the price at the beginning of the period of interest (here a month).
Close refers to the "closing price, i.e., to the price at the end of the period of interest. You can check for yourself that the closing price at the end of a month is indeed very close to the opening price at the beginning of the next month.
High and **Low** refer to the highest and the lowest price over the period of interest.
Volume refers to the trading volume during the period of interest.
Finally, **Adjusted Close** refers to the price adjusted for dividend distribution (and split) in the following time period of interest. This series is designed for calculating investment returns.

Calculate returns

For this you should use the "adjusted close" price. Insert a new column to the right of the "adjusted close" column. Calculate the return from one period to the next by dividing the adjusted close in the current period, by the adjusted close in the previous period, subtracting one. Change the format of the cell so that everything is displayed in percentage terms.

Calculate the mean and standard deviation

For both the stock and the SP500, calculate the mean return and the standard deviation of return. The Excel function for the mean is AVERAGE, and the Excel function for the standard deviation is STDEVP.

Which one is the most variable, the stock or the SP500? Why should we expect the SP500 to be less variable than the stock?

Correlation

Calculate the correlation coefficient between the two returns series (Excel function is CORREL). What is the sign of the correlation? Is it what you expected and why?

Scatter plot

Produce a scatter plot of the stock return against the SP500 return. How does the sign of the correlation appear on the scatter plot?

Exercise 2: do daily returns follow a normal distribution?

Download data

Go to yahoo.com and download daily price data for the SP500: from January 1st, 2006 to January 1st, 2022. The stock symbol fro SP500 is "^GSPC". Create return data using the adjusted close price.

Plot the level of the SP500 over the time period

Briefly describe the broad historical evolution of the SP500 since 2006.

Calculate means and standard deviation of the return

Use the same Excel functions as in Exercise 1.

Create a frequency table for the data

Create a frequency table for this data using the Excel function "Frequency", explained at the end of this handout. Use the following bins: -10,-9.5,-9,-8.5,-8, -7.5, -7 etc… 7, 7.5, 8, 8.5, 9,9.5,10.

Create a similar frequency table for the normal distribution

Create a frequency table assuming that the data is generated by a normal distribution. That is, for each bin, calculate the probability of falling in that bin according to a normal distribution with the same mean and standard deviation as calculated above.

How do the two distributions differ?

Provide numerical answer to the following questions:

What is the ratio of the true probability and of the normal probability of observing a daily return less than -5%?

Harder question: What is the true probability and the normal probability of observing **at least one daily return of less than -5%** over a year? To answer this question, assume that returns are independently and identically distributed over the year.

From your calculations, do you conclude that the two distributions assign similar probabilities to extreme events?

Financial firm often seek to assess the probability of making large losses on their portfolio (the technical concept that is used is VaR or "Value at Risk). See http://en.wikipedia.org/wiki/Value_at_risk for some readings. According to the above, is the normal distribution appropriate for a Value at Risk calculation?

Bitcoin returns

On Yahoo finance, download daily data on the Bitcoin-to-USD exchange rate (the symbol is BTC-USD), using the maximum time period available (starting in 2014, ending now). Briefly describe the evolution of the exchange rate over the time period. Calculate the mean return, and the standard deviation of return. How does the empirical distribution of Bitcoin returns differ from a normal distribution?

Notes on two Excel formulas

The FREQUENCY function

Here is an example of the "Frequency function." Consider the following table:

	H
1	-1
2	2
3	-3
4	4
5	0
6	0
7	3
8	1
9	-0.5
10	2

The "frequency" function allows you to calculate the number of observations below a certain threshold. For instance, if you enter:

FREQUENCY($H$1:$H$10,4), you obtain the number of observation that, in the column $H$1:$H$10, are less than 4.

To calculate the number of observation between 3 and 4, you simply need to calculate the number of observation less than 4, and subtract the number of observations less than 3.

The normal distribution in Excel

For any given mean and standard deviation and any given threshold, you can use Excel to calculate the probability of falling below the threshold, according to the normal distribution. That is, you can calculate the area below the "bell curve" that is to the left of the threshold of interest.

For instance, if the mean is 0.1, the standard deviation is 3, and the threshold is 4, then the formula for the probability is:

NORM.DIST(4,0.1,3,1)

The firms argument is the threshold, the second argument is the mean, and the third argument is the standard deviation. The last argument stand for "cumulative" (i.e., for the fact that we calculate the probability up to the threshold).

To calculate the probability of falling in between two thresholds, say 4 and 5, the formula is:

NORM.DIST(5,0.1,3,1) - NORM.DIST(4,0.1,3,1)

That is, the probability of falling below 5, minus the probability of falling below 4.