**Data Project 3:**

***Total Possible Points: 6+3+3=12 pts***

The data file “stock\_price.csv” contains the historical daily rate of return for two stocks: Bank of America (***BAC***) and JP Morgan (***JPM***) between the period of 2016 and 2018

**Part 1**

The goal of **Part 1** in data project 3 is to perform sampling methods learnt in the course on this dataset. Unless specified, students are expected to use R to perform the analysis and answer questions. To obtain full points, you need to show both the R code and the corresponding result.

**Q1)** Use R command to import the datafile into R and name this data as **stock**. Use R head() function to preview the first 10 observations in R. (**1 pts**)

**Q2)** Follow the procedure below to create a simple random sample for the stock data: (**2 pts**)

2.1) Use nrow() function on the stock data to obtain the total number of observations for the data

2.2) Use sample() function to generate 100 *non-duplicate* integers between 1 and the total # of obs of the data, save this 100 random integer as row\_index in R.

2.3) Use the row\_index above to select a subset of the data, and save it as **SRS\_100** in R

2.4) Apply the summary() function on **SRS\_100** data to summarize the daily return of return for BAC and JPM

**Q3)** Follow the procedure below to perform a stratified random sampling for the stock data: (**2.5 pts**)

3.1) Based on the year variable, use subset() function or which() function, or any other method to split the **stock** data into 3 stratum: **stock2016**, **stock2017** and **stock2018**, which represents the stock return in 2016, 2017 and 2018.

3.2) Apply the summary() function on each stratum dataset to summarize the stock return over different period.

3.3) Apply the simple random sampling procedure in Q2) for each stratum. Select 30 observations from **stock2016**, 30 observations from **stock2017** and 40 observations from **stock2018**; save each SRS as **SRS2016\_30**, **SRS2017\_30** and **SRS2018\_30**

3.4) Use rbind() function to combine **SRS2016\_30**, **SRS2017\_30** and **SRS2018\_30,** save the combined data as **StratifiedRS\_100.** Usesummary()function to summarize this  **StratifiedRS\_100** data.

**Part 2**

The goal of **Part 2** in data project 3 is to use the knowledge of probability rules to examine the return pattern between the two stocks. Unless specified, students are expected to use R to perform the analysis and answer questions. To obtain full points, you need to show both the R code and the corresponding result.

**Q1)** Use ifelse() function to generate two categorical variables ***BAC\_Profit*** and ***JPM\_Profit*** that represents the events whether trader can make daily profit on that day from each stock, i.e. return > 0. There are two categories for both variables: 0 and 1.

If the return of BAC is greater than 0, then ***BAC\_Profit*** takes the value 1, otherwise it takes the value 0. (**0.5 pts**)

**Q2)** Based on the Event variable ***BAC\_Profit*** and ***JPM\_Profit****,*calculate the following probabilities:

* unconditional probability of a trader makes profit on BAC. (**0.5 pts**)
* unconditional probability of a trader makes profit on JPM. (**0.5 pts**)
* joint probability of making profit on both BAC and JPM (**0.5 pts**)
* conditional probability of BAC makes profit, given that JPM makes profit (**0.5 pts**)

Discuss whether the two events making profit on BAC and making profit on JPM are independent events (**0.5 pts**)

**Part 3**

The goal of **Part 3** in data project 3 is to use the knowledge of random variables to examine the distribution and perform the corresponding calculation. In this question, ***student can use either R or Excel*** to perform the calculation. To obtain full points, you need to provide the complete table/code/function to show your calculation procedure. **Part 3** questions are not based on any dataset.

**Q1)** Assume that your job is to make forecast on the stock performance, and your bonus is tied to the accuracy of your forecast. The table below shows the probability of making satisfying prediction and the corresponding bonus. Evaluate the expected value and standard deviation of your bonus based on this table. (**1 pts**)

|  |  |  |
| --- | --- | --- |
| **Probability** | **Forecast Error** | **Bonus** |
| 0.10 | > 50% | 0 |
| 0.30 | between 20% and 50% | 1000 |
| 0.35 | between 5% and 20% | 5000 |
| 0.25 | < 5% | 20000 |

**Q2**) Assume that you want to create a portfolio using two stocks A and B. The table below shows the performance of each individual stock and the correlation between them.

|  |  |  |
| --- | --- | --- |
|  | **Stock A** | **Stock B** |
| **Expected Return** | 15.00% | 8.00% |
| **Standard Deviation** | 17.00% | 5.00% |
| **Correlation r** | -0.3 | |

Calculate the performance (expected return and variance) for the following three portfolios. (**2 pts**)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Weigh of A** | **Weight of B** | **Portfolio Return** | **Portfolio Variance** |
| Portfolio 1 | 10% | 90% |  |  |
| Portfolio 2 | 30% | 70% |  |  |
| Portfolio 3 | 50% | 50% |  |  |
| Portfolio 4 | 70% | 30% |  |  |