Case Study 2

Question 01 (10 points)

Create the X matrix and print it from SAS, R, and Python.

$$X = \begin{pmatrix} 4 & 5 & 1 & 2 \\ 1 & 0 & 3 & 5 \\ 2 & 1 & 8 & 2 \end{pmatrix}$$
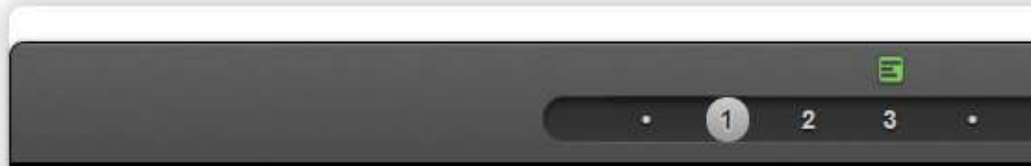
SAS code

R code

Python Code

Question 02 (15 points)

Week 11 Lecture Assignment

Unit 11 | Modeling Financial Data

Please watch videos1 and 2 in week 11 lecture assignment. You can download the code which used for S&P from files tab.

Please do the following with your assigned stock.
- Download the data.
- Calculate log returns.
- Calculate volatility measure.
- Calculate volatility over entire length of series for various three different decay factors.
- Plot the results, overlaying the volatility curves on the data, just as was done in the S&P example.

Group	Stock
Yao and Robert	AGIO
Max and Josh	ADP
Patrick and Sunna	GWPH
Robert and Jason	TCP
Simrat and Priyanka	CO
Samira and Ghizlane	XOM

If you cannot download data from your assigned stock you can pick a different stock from the following list.

PM, PG, HON, URTY, PEN

You can use Yahoo finance or google finance to find your data!
Submit your final code and the plot.

Question 03 (20 points)

The built-in data set called *Orange* in R is about the growth of orange trees. The Orange data frame has 3 columns of records of the growth of orange trees.

Variable description
Tree : an ordered factor indicating the tree on which the measurement is made. The ordering is according to increasing maximum diameter.

age : a numeric vector giving the age of the tree (days since 1968/12/31)
circumference : a numeric vector of trunk circumferences (mm). This is probably "circumference at breast height", a standard measurement in forestry.

a) Calculate the mean and the median of the trunk circumferences for different size of the trees. (Tree)
b) Make a scatter plot of the trunk circumferences against the age of the tree. Use different plotting symbols for different size of trees.
c) Display the trunk circumferences on a comparative boxplot against tree. Be sure you order the boxplots in the increasing order of maximum diameter.

Submit your final R code and necessary plots for each part.

Question 04 (45 points)

Download "Temp" data set (check your SMU email)

(i)   Find the difference between the maximum and the minimum monthly average temperatures for each country and report/visualize top 20 countries with the maximum differences for the period **since 1900**.

(ii)   Select a subset of data called "UStemp" where US land temperatures from 01/01/1990 in Temp data. Use UStemp dataset to answer the followings.

        a) Create a new column to display the monthly average land temperatures in Fahrenheit (°F).

        b) Calculate average land temperature by year and plot it. The original file has the average land temperature by month.

        c) Calculate the one year difference of average land temperature by year and provide the maximum difference (value) with corresponding years.

        (for example, year 2000: add all 12 monthly averages and divide by 12 to get average temperature in 2000. You can do the same thing for all the available years. Then you can calculate the one year difference as 1991-1990, 1992-1991, etc)

(iii)   Download "CityTemp" data set (check your SMU email). Find the difference between the maximum and the minimum temperatures for each major city and report/visualize top 20 cities with maximum differences for the period since 1900.

(iv)   Compare the two graphs in (i) and (iii) and comment it.

You can use either R or Python

Please submit your final code with necessary plots for each part!

Question 05 (10 points)

Semester End Bonus

Run the following code in R and paste the final graph.

```
t<-seq(0,10,length=1000)
x<-sqrt(t)*cos(2*pi*t)
y<-sqrt(t)*sin(2*pi*t)
plot(x,y,axes=F,type="l",lwd=3,xlab="x(t)",ylab="y(t)",col="red")

axis(1,at=seq(-3,3,by=0.5),labels=seq(-3,3,by=0.5))
axis(2)
box()
title(main=expression(
paste("(x(t),y(t)) with polar coordinates",
(list(sqrt(t),2*pi*t))
)))
```

Thank you for your effort throughout this semester!

I wish good luck for our future Data Scientists!!