**Student: Seif Kungulio**

**Date: 01/25/2025**

**Subject: Project 2**

**Class: DSCI 502**

**Section: 01W**

**Instructor: Sean Yang**

**File Name: Project2\_Kungulio\_Seif.docx**

1. Read the dataset in SP500.xls into R. Call the loaded data SP500. Make sure that you have the directory set to the correct location for the data.

A screenshot of a computer

Description automatically generated

|  |
| --- |
| > # 1. Read the dataset in SP500.xls into R. Call the loaded data SP500.  > # Make sure that you have the directory set to the correct location  > # for the data.  >  > # Set the working directory  > setwd("C:/Projects/DSCI 502/Week 2")  >  > # Import necessary libraries  > library(readxl)  >  > # Import the data set  > SP500 <- read\_excel("SP500.xls")  >  > # Display the dimension of the dataframe  > dim(SP500)  [1] 1773 6  >  > # Verify that the file was loaded correctly  > str(SP500)  tibble [1,773 × 6] (S3: tbl\_df/tbl/data.frame)  $ Date : num [1:1773] 1871 1871 1871 1871 1871 ...  $ SP500 : num [1:1773] 4.44 4.5 4.61 4.74 4.86 4.82 4.73 4.79 4.84 4.59 ...  $ Dividend: num [1:1773] 0.26 0.26 0.26 0.26 0.26 0.26 0.26 0.26 0.26 0.26 ...  $ Earnings: num [1:1773] 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 0.4 ...  $ CPI : num [1:1773] 12.5 12.8 13 12.6 12.3 ...  $ Rate : num [1:1773] 5.32 5.32 5.33 5.33 5.33 ... |
|  |
|  |

1. How many rows are in the data set? How many columns are in the data set?

A screenshot of a computer

Description automatically generated

|  |
| --- |
| > # 2. How many rows are in the data set? How many columns are in the data set?  >  > # Display the number of rows  > cat("There are", nrow(SP500), "number of rows\n")  There are 1773 number of rows  >  > # Display the number of columns  > cat("There are", ncol(SP500), "number of columns\n")  There are 6 number of columns |
|  |
| |  | | --- | |  | |

1. Select the following three columns: SP500, CPI, and Rate.

A screenshot of a computer

Description automatically generated

|  |
| --- |
| > # 3. Select the following three columns: SP500, CPI, and Rate.  >  > # Create a variable to hold the selected three columns  > three\_cols.df <- SP500[,c("SP500", "CPI", "Rate")]  >  > # Print the 'three\_cols.df' dataframe showing the first six rows  > head(three\_cols.df)  # A tibble: 6 × 3  SP500 CPI Rate  *<dbl>* *<dbl>* *<dbl>*  1 4.44 12.5 5.32  2 4.5 12.8 5.32  3 4.61 13.0 5.33  4 4.74 12.6 5.33  5 4.86 12.3 5.33  6 4.82 12.1 5.34 |
|  |
|  |

1. Select the 10th, 100th, 500th, and 1500th rows.

A screenshot of a computer

Description automatically generated

|  |
| --- |
| > # 4. Select the 10th, 100th, 500th, and 1500th rows.  >  > # Create a variable to hold the selected four rows  > four\_rows.df <- SP500[c(10, 100, 500, 1500), ]  >  > # Print the 'four\_rows.df' dataframe  > print(four\_rows.df)  # A tibble: 4 × 6  Date SP500 Dividend Earnings CPI Rate  *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*  1 1871. 4.59 0.26 0.4 12.4 5.35  2 1879. 3.77 0.187 0.333 8.18 4.17  3 1912. 9.81 0.477 0.663 9.70 4.27  4 1995. 615. 13.8 34.0 154. 5.71 |
|  |
|  |

1. Select all the observations such that SP500 is greater than 2000 or CPI is less than 100.

A screenshot of a computer

Description automatically generated

|  |
| --- |
| > # 5. Select all the observations such that SP500 is greater than 2000 or  > # CPI is less than 100.  >  > # Filter data according the prescribed condition  > obs1.df <- SP500[(SP500$SP500 > 2000) | (SP500$CPI < 100), ]  >  > dim(obs1.df) # Check the dimension of the resultant dataframe  [1] 1395 6  > head(obs1.df) # Display the first six rows of the dataframe  # A tibble: 6 × 6  Date SP500 Dividend Earnings CPI Rate  *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*  1 1871. 4.44 0.26 0.4 12.5 5.32  2 1871. 4.5 0.26 0.4 12.8 5.32  3 1871. 4.61 0.26 0.4 13.0 5.33  4 1871. 4.74 0.26 0.4 12.6 5.33  5 1871. 4.86 0.26 0.4 12.3 5.33  6 1871. 4.82 0.26 0.4 12.1 5.34  > tail(obs1.df) # Display the last six rows of the dataframe  # A tibble: 6 × 6  Date SP500 Dividend Earnings CPI Rate  *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*  1 2018. 2654. 50.3 118. 251. 2.87  2 2018. 2701. 50.7 120. 252. 2.98  3 2018. 2754. 51.0 122. 252. 2.91  4 2018. 2794. 51.4 125. 252. 2.89  5 2018. 2858. 51.9 128. 252. 2.89  6 2018. 2902. 52.3 130. 252. 3 |
|  |

1. Select the data such that Earnings greater than 50 and Rate less than 3 with columns SP500 and Dividend only.

A screenshot of a computer

Description automatically generated

|  |
| --- |
| > # 6. Select the data such that Earnings greater than 50 and Rate  > # less than 3 with columns SP500 and Dividend only.  >  > # Filter data according the prescribed condition  > obs2.df <- SP500[(SP500$Earnings > 50) & (SP500$Rate < 3), ]  > #View(obs2.df) # View the dataframe in another tab  >  > # Filter to return the required columns only  > obs2.df <- obs2.df[, c("SP500", "Dividend")]  >  > dim(obs2.df) # Check the dimension of the resultant dataframe  [1] 89 2  > head(obs2.df) # Display the first six rows of the dataframe  # A tibble: 6 × 2  SP500 Dividend  *<dbl>* *<dbl>*  1 1087. 22.2  2 1122. 22.4  3 1172. 22.5  4 1199. 22.6  5 1185. 24.9  6 1174. 25.2 |
|  |
|  |

1. Remove the entire column of Rate.

A screenshot of a computer program

Description automatically generated

|  |
| --- |
| > # 7. Remove the entire column of Rate.  >  > # Create a variable to hold the dataframe  > without\_rate.df <- SP500[, -6]  >  > # Display the first six rows  > head(without\_rate.df)  # A tibble: 6 × 5  Date SP500 Dividend Earnings CPI  *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*  1 1871. 4.44 0.26 0.4 12.5  2 1871. 4.5 0.26 0.4 12.8  3 1871. 4.61 0.26 0.4 13.0  4 1871. 4.74 0.26 0.4 12.6  5 1871. 4.86 0.26 0.4 12.3  6 1871. 4.82 0.26 0.4 12.1 |
|  |
|  |

1. Real price is the inflation adjusted price, which is given by the following formula: Real Price at time t = (SP 500 price at time t)\* CPI(t)/CPI(2018.09), where CPI(2018.09) is the latest consumer price index in the data set. Based on formula above, you need to add one more column, RealPrice.

A screenshot of a computer

Description automatically generated

|  |
| --- |
| > # 8. Real price is the inflation adjusted price, which is given by  > # the following formula:  > # Real Price at time t = (SP 500 price at time t)\* CPI(t)/CPI(2018.09),  > # where CPI(2018.09) is the latest consumer price index in the data set.  > # Based on formula above, you need to add one more column, RealPrice.  >  > # Extract the latest CPI using tail() function  > latest\_CPI <- tail(SP500$CPI, n = 1)  > #print(latest\_CPI)  >  > # Calculate Real Price and assign to RealPrice column  > SP500$RealPrice <- SP500$SP500 \* SP500$CPI / latest\_CPI  >  > head(SP500) # Display the first six rows of the dataframe  # A tibble: 6 × 7  Date SP500 Dividend Earnings CPI Rate RealPrice  *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*  1 1871. 4.44 0.26 0.4 12.5 5.32 0.219  2 1871. 4.5 0.26 0.4 12.8 5.32 0.229  3 1871. 4.61 0.26 0.4 13.0 5.33 0.238  4 1871. 4.74 0.26 0.4 12.6 5.33 0.236  5 1871. 4.86 0.26 0.4 12.3 5.33 0.236  6 1871. 4.82 0.26 0.4 12.1 5.34 0.231 |
|  |
|  |

1. Real earnings are the inflation adjusted earnings, which is given by the following formula:

Real earnings at time t = (earnings at time t)\* CPI(t)/CPI(2018.09), where CPI(2018.09) is the latest consumer price index in the data set.

Based on formula above, you need to add one more column, RealEarnings.

A screenshot of a computer screen

Description automatically generated

|  |
| --- |
| > # 9. Real earnings are the inflation adjusted earnings, which is given by  > # the following formula:  > # Real earnings at time t = (earnings at time t)\* CPI(t)/CPI(2018.09),  > # where CPI(2018.09) is the latest consumer price index in the data set.  > # Based on formula above, you need to add one more column, RealEarnings.  >  > # Calculate Real Earnings and assign to RealEarnings column  > SP500$RealEarnings <- SP500$Earnings \* SP500$CPI / latest\_CPI  >  > head(SP500) # Display the first six rows of the dataframe  # A tibble: 6 × 8  Date SP500 Dividend Earnings CPI Rate RealPrice RealEarnings  *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*  1 1871. 4.44 0.26 0.4 12.5 5.32 0.219 0.0197  2 1871. 4.5 0.26 0.4 12.8 5.32 0.229 0.0204  3 1871. 4.61 0.26 0.4 13.0 5.33 0.238 0.0207  4 1871. 4.74 0.26 0.4 12.6 5.33 0.236 0.0199  5 1871. 4.86 0.26 0.4 12.3 5.33 0.236 0.0194  6 1871. 4.82 0.26 0.4 12.1 5.34 0.231 0.0191 |
|  |
|  |

1. Price to earnings ratio is given by the following formula:

P/E Ratio = RealPrice/RealEarnings

Based on the formula above, please add one more column, PERatio.

A screenshot of a computer screen

Description automatically generated

|  |
| --- |
| > # 10. Price to earnings ratio is given by the following formula:  > # P/E Ratio = RealPrice/RealEarnings  > # Based on the formula above, please add one more column, PERatio.  >  > # Calculate P/E Ratio and assign to PERatio column  > SP500$PERatio <- SP500$RealPrice / SP500$RealEarnings  >  > head(SP500) # Display the first six rows of the dataframe  # A tibble: 6 × 9  Date SP500 Dividend Earnings CPI Rate RealPrice RealEarnings PERatio  *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>* *<dbl>*  1 1871. 4.44 0.26 0.4 12.5 5.32 0.219 0.0197 11.1  2 1871. 4.5 0.26 0.4 12.8 5.32 0.229 0.0204 11.2  3 1871. 4.61 0.26 0.4 13.0 5.33 0.238 0.0207 11.5  4 1871. 4.74 0.26 0.4 12.6 5.33 0.236 0.0199 11.9  5 1871. 4.86 0.26 0.4 12.3 5.33 0.236 0.0194 12.1  6 1871. 4.82 0.26 0.4 12.1 5.34 0.231 0.0191 12.0 |
|  |
|  |