HW2 Part 2

**Question 1)**

**For part A and B**

PlantGrowth is a dataset in R that contains crop weights of a control group and two treatment groups.

*#Code to Get Data*

*library(datasets)*

*data(PlantGrowth)*

Perform the following operations: Create two separate datasets: dataset1 contains datapoints of treatment 1 group and control group, and dataset2 contains datapoints of treatment 2 group with the control group.

**A)** Now compute the difference estimator for each dataset that were just created. (treatment 1 vs control group for dataset1, and treatment 2 vs control group for dataset2)

**B)** From the PlantGrowth dataset what is the average crop weight of the control group, treatment 1 group, and treatment 2 group, comment on which group has the highest average?

**Use the dataset “Min\_Wage.csv” and the information below to answer part C, D and E.**

The Minimum Wage Law protects the right of workers to get a minimum wage. Consider a scenario where the law of minimum wage was changed just in the state of New Jersey (i.e., law has not been changed in other states). We want to use the data from company XYZ to observe the difference in hours worked by full time employees in New Jersey before and after the law was changed.

Note: The variable 'State’ indicates the citizenship of the worker, i.e., State = New Jersey ,the worker is from NJ, else worker is not from NJ (is from Philadelphia).

Note: The variable fte contains the number of hours worked by full time employee.

Note: The variable d indicates whether or not the data was collected before or after the law changed, i.e. d = 1 indicates the data was collected after the law was changed, and d = 0 indicates the data was collected before the law was changed.

**C)** In the above problem, classify the workers into four groups and assign the corresponding group with the group title (A,B,C and D) where the group titles are as follows:

|  |  |  |
| --- | --- | --- |
|  | Before | After |
| Control | A | C |
| Treated | B | D |

**Note:** This is a theoretical question, you just need to identify each group with a short description of what the group is. You **don’t** need to create a column categorizing the datapoints into groups.

**D)** To estimate the difference in difference we need four averages for the above categorized groups i.e., control group before change, control group after change, treatment group before change and treatment group after change. Compute the following

(i) Calculate the mean of the ‘fte’ variable for each of the four groups in R and print them

(ii) Using these averages estimate the value of the difference in difference

**E)** Estimate the DID (Difference in Difference) using a regression model

**Answers Q1:**

**A):**   
The b1 coefficient for each model is -0.37 and 0.49 respectively. See summaries below.

*library(datasets)*

*library(dplyr)*

*data(PlantGrowth)*

*# create two datasets*

*df1 <- PlantGrowth %>% filter(group=="ctrl"|group=="trt1")  
df2 <- PlantGrowth %>% filter(group=="ctrl"|group=="trt2")*

*# fit linear regression*

*Model1 <- lm(weight ~ ., data = df1)*

*summary(model1)*

*Model2 <- lm(weight ~ ., data = df2)*

*summary(model2)*

**B):**

**5.03,4.66 and 5.53 and Treatment 2 group has the highest average**

**Explanation:** The intercept (5.03) is the average crop weight of the control group. Treatment Group 1 has diff estimator of -0.37 which makes the average weight for that group 4.66, and Treatment Group 2 has diff estimator of 0.49 which makes the average weight for that group 5.53. See summary below.

**C):**

A - Worker’s from Philadelphia before the law was changed.

B - Worker’s from New Jersey before the law was changed.

C - Worker’s from Philadelphia after the law was changed.

D - Worker’s from New Jersey after the law was changed.

**D):**

*min\_wage <- read.csv("Min\_Wage.csv")*

*min\_wage<- min\_wage%>%*

*mutate(nj = ifelse(State == "New Jersey",1,0))*

*min\_wage $d = as.numeric(min\_wage $d)*

*min\_wage $fte = as.numeric(min\_wage $fte)*

*# The active period of Law change is indicated by variable 'd'*

*# The citizenship of New Jersey is indicated by variable 'nj'*

*a = sapply(subset(min\_wage, nj == 0 & d == 0, select=fte), mean)*

*#a =* 23.38

*b = sapply(subset(min\_wage, nj == 1 & d == 0, select=fte), mean)*

*#b =* 20.43058

*c = sapply(subset(min\_wage, nj == 0 & d == 1, select=fte), mean)*

*#c =* 21.09667

*d = sapply(subset(min\_wage, nj == 1 & d == 1, select=fte),mean)*

*#d =* 20.89725

*# Difference in difference parameter*

*DID = (d-b)-(c-a)*

*print(DID) #Answer = 2.75*

**E):**

*model = lm(fte ~ nj+ d + nj\*d,data = min\_wage)*

*summary(model)*

*#Difference in difference estimator is the coefficient of the interaction term which is 2.75*

Rubric for Peer Assessment

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Question No.** | **Total Pts.** | **Ratings** | | |
| 1A | 5 | Marks: 5    Fitted it as two models (1 mark for each model) and got the answers correct (1.5 for each right answer) | Marks: 3    Got the right answers but did not split it into two datasets | Marks: 0    Incorrect answer |
| 1B | 6 | Marks: 6  1.5 mark for the code (Generic as the mean can be computed using ‘sapply’ as well)  1.5 marks for each right average value |  | Marks: 0    Incorrect answer. |
| 1C | 8 | Marks: 8    2 marks each for each category  assigned correctly. |  | Marks: 0    Incorrect answer. |
| 1D | 7 | Marks: 7    Find all the four average values (1.5 mark for each correct average),  Find the DID (1 marks) |  | Marks: 0    Incorrect answer |
| 1E | 4 | Marks: 4    Fit the correct regression model (2 marks), identify the DID as the interaction term (2 marks). |  | Marks: 0    Incorrect answer |

**Question 2)**

For the following questions, use the dataset Berkshire.csv with the following variables:

* Column (1): Date, Calendar Date
* Column (2): BRKret, Berkshire Hathaway’s monthly return
* Column (3): MKT, the return on the aggregate stock market
* Column (4): RF, the risk free rate of return

You may/may not need the following dependencies:

“PerformanceAnalytics” package

“lubridate” package

Return data is in decimal format, 0.01=1%.  **Round all answers to the nearest hundredth, for example 12.34%.**

1. Find the standard deviation of Berkshire Hathaway over the sample period. (1 point)
2. Find Berkshire Hathaway’s average return over the sample period? Use arithmetic average (1 point)
3. By what percentage per month on average has Berkshire Hathaway outperformed the market? (1 point)
4. $10,000 invested in Berkshire Hathaway at the start of the sample period would have GROWN TO \_\_\_\_ by the end of the sample period. (1 point)

(Hint: make sure to calculate how much the investment would have grown to, not how much it would grow by)

1. Plot the cumulative return of Berkshire and Market across all years and include a legend. Describe your observation. (1 point)

**Answer Q2**

**A)** 6.75%   
*round(sd(data$BrkRet)\*100,2)*

**B)** 1.90%

*round(mean(data$BrkRet)\*100,2)*

**C)** 0.88%   
*round(mean(data$BrkRet)\*100,2) – round(mean(data$MKT)\*100,2)*

**D)** $41,449,896

#*converting dates to standard YYYY-MM-DD format   
data$Date <- mdy(data$Date)  
  
#Sorting data by dates   
data2<- data[order(data$Date),]   
  
#create an xts dataset   
All.dat <- xts(data2[,-1], order.by = data2$Date)   
  
#Calculate Compound Return for the fund across all the data  
Return.cumulative(All.dat$BrkRet,geometric = TRUE)   
cum\_ret <- Return.cumulative(All.dat$BrkRet,geometric = TRUE)[1]   
  
#Cumulative Returns chart over time   
#Check chart in Plots Tab on bottom right in R Studio  
chart.CumReturns(All.dat$BrkRet,wealth.index = FALSE, geometric = TRUE)*

*10000 \* (1+cum\_ret)*

**E)**   
*dat2 <- All.dat[ , 1:2]   
chart.CumReturns(dat2,wealth.index = FALSE, geometric = TRUE, legend.loc = "topleft")*

Berkshire Hathaway’s return was significantly more than the market.   
If you were to invest $1 at the beginning of period, your investment would grow to approximately $4000 ($4143.99 to be exact) compared to the market return of approximately $100 ($96.22)

*#Return.cumulative(All.dat$MKT,geometric = TRUE)*

*#96.22*

*#Return.cumulative(All.dat$BrkRet,geometric = TRUE)*

*#4143.99*

|  |  |  |  |
| --- | --- | --- | --- |
| **Question No.** | **Total Pts.** | **Ratings** | |
| 2A | 2 | Marks: 2  Correct Answer  Marks: 1  Partial Credit | Marks: 0    Incorrect answer |
| 2B | 2 | Marks: 2  Correct Answer  Marks: 1  Partial Credit | Marks: 0    Incorrect answer |
| 2C | 2 | Marks: 2  Correct Answer  Marks: 1  Partial Credit | Marks: 0    Incorrect answer |
| 2D | 2 | Marks: 2  Correct Answer  Marks: 1  Partial Credit | Marks: 0    Incorrect answer |
| 2E | 2 | Marks: 2  Correct Answer  Marks: 1  Partial Credit | Marks: 0    Incorrect answer |