Homework 2 – Part2

Q1.) 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:

(i) Create two separate datasets, one with datapoints of treatment 1 group along with control group and other with datapoints of treatment 2 group with the control group.

A)

Now compute the difference estimator for treatment 1 and treatment 2 datasets that were created, in comparison with the control group?

Explanation: The b_1 coefficient for each model is -0.37 and 0.49 respectively. See summaries below.

```
> model_tr1 = lm(weight ~ group, data = data_tr1)
> summary(model_tr1)

Call:
lm(formula = weight ~ group, data = data_tr1)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0710 -0.4938  0.0685  0.2462  1.3690

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.0320     0.2202  22.850 9.55e-15 ***
grouptrt1    -0.3710     0.3114  -1.191    0.249
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6964 on 18 degrees of freedom
Multiple R-squared:  0.07308, Adjusted R-squared:  0.02158
F-statistic: 1.419 on 1 and 18 DF,  p-value: 0.249

> model_tr2 = lm(weight ~ group, data = data_tr2)
> summary(model_tr2)

Call:
lm(formula = weight ~ group, data = data_tr2)

Residuals:
   Min     1Q Median     3Q    Max
-0.862 -0.410 -0.006  0.280  1.078

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.0320     0.1637  30.742   <2e-16 ***
grouptrt2     0.4940     0.2315   2.134   0.0469 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.5176 on 18 degrees of freedom
Multiple R-squared:  0.2019, Adjusted R-squared:  0.1576
F-statistic: 4.554 on 1 and 18 DF,  p-value: 0.04685
```

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?

Solution: 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.

```
> model_tr1 = lm(weight ~ group, data = data_tr1)
> summary(model_tr1)

Call:
lm(formula = weight ~ group, data = data_tr1)

Residuals:
    Min      1Q  Median      3Q     Max
-1.0710 -0.4938  0.0685  0.2462  1.3690

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.0320     0.2202  22.850 9.55e-15 ***
grouptrt1    -0.3710     0.3114  -1.191    0.249
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.6964 on 18 degrees of freedom
Multiple R-squared:  0.07308, Adjusted R-squared:  0.02158
F-statistic: 1.419 on 1 and 18 DF,  p-value: 0.249

> model_tr2 = lm(weight ~ group, data = data_tr2)
> summary(model_tr2)

Call:
lm(formula = weight ~ group, data = data_tr2)

Residuals:
   Min     1Q Median     3Q    Max
-0.862 -0.410 -0.006  0.280  1.078

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   5.0320     0.1637  30.742   <2e-16 ***
grouptrt2     0.4940     0.2315   2.134   0.0469 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5176 on 18 degrees of freedom
Multiple R-squared:  0.2019, Adjusted R-squared:  0.1576
F-statistic: 4.554 on 1 and 18 DF,  p-value: 0.04685
```

Page Break

Use the following for 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 an 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) (i.e., control group before change to the group A etc.). where the group titles are as follows:

	Before	After
Control	A	C
Treated	B	D

Solution:

A - Worker's from Philadelphia who are inactive after the law was changed.

B - Worker's from New Jersey who are inactive after the law was changed.

C - Worker's from Philadelphia who are active after the law is changed.

D - Worker's from New Jersey who are active after the law is changed.

Part 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

Solution

```{r}
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)
b = sapply(subset(min_wage, nj == 1 & d == 0, select=fte), mean)
c = sapply(subset(min_wage, nj == 0 & d == 1, select=fte), mean)
d = sapply(subset(min_wage, nj == 1 & d == 1, select=fte),mean)
# Difference in difference parameter
DID = (d-b)-(c-a)
print(DID)  #Answer = 2.75
```

E) Estimate the DID (Difference in Difference) using regression model.

Solution # Using Regression

model = lm(fte ~ nj+ d + nj*d,data = min_wage)

summary(model)

```
Call:
lm(formula = fte ~ nj + d + nj * d, data = min_wage)

Residuals:
    Min      1Q  Median      3Q     Max
-21.097  -6.472  -0.931   4.603  64.569

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   23.380      1.098  21.288   <2e-16 ***
nj            -2.949      1.224  -2.409   0.0162 *
d             -2.283      1.553  -1.470   0.1419
nj:d           2.750      1.731   1.588   0.1126
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 9.511 on 764 degrees of freedom
Multiple R-squared:  0.007587,  Adjusted R-squared:  0.00369
F-statistic: 1.947 on 3 and 764 DF,  p-value: 0.1206
```

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

Q2.)

For the following 2 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

Round all answers to the nearest hundredth.

a. Find the standard deviation of Berkshire Hathaway over the sample period. (1 point)

Answer: 6.75%

Explanation: round(sd(data$BrkRet)*100,2)

b. Find Berkshire Hathaway’s average return over the sample period? (1 point)

Answer: 1.90%

Explanation: round(mean(data$BrkRet)*100,2)

c. By what percentage per month on average has Berkshire Hathaway outperformed the market? (1 point)

Answer: 0.88%

Explanation: round(mean(data$BrkRet)*100,2) – round(mean(data$MKT)*100,2)

d. $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)

Answer: $41,449,896

Explanation:

#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)
```

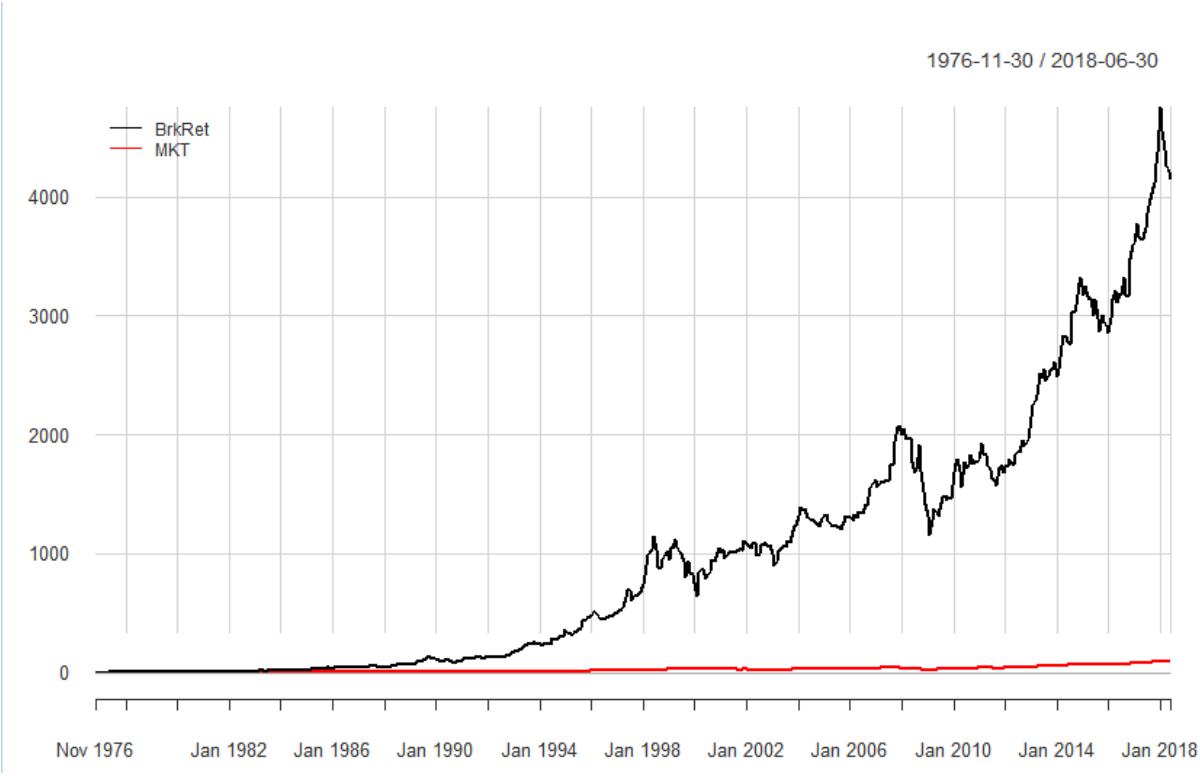
e. Plot the cumulative return of Berkshire and Market across all years and include a legend. Describe your observation. (1 point)

```
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 2 (4 points)

f. What is Berkshire Hathaway's monthly Sharpe ratio? (1 point)
Answer: 0.23
Explanation: SharpeRatio(All.dat$BrkRet, All.dat$RF)

g. What is the Sharpe Ratio for the market index? Comparing this value to Berkshire Hathaway's Sharpe ratio. Which one is higher and what does that mean? (2 points)

Answer: 0.15
Berkshire Hathaway has a higher Sharpe ratio which indicates higher reward per unit risk.
Explanation: SharpeRatio(All.dat$MKT, All.dat$RF)

h. What is Berkshire Hathaway's estimated beta? (1 point)
Answer: 0.70
Explanation: summary(lm(All.dat$BrkRet ~ All.dat$MKT))

i. On a monthly basis, what is Jensen's alpha for Berkshire Hathaway? (1 point)
Answer: 1.08%
All.dat <- transform(All.dat, MktExcess= MKT-RF, FundExcess = BrkRet-RF)
Alpha = lm(FundExcess ~ MktExcess, data = All.dat)
summary(Alpha)