Yida Xu HW2

Exercise1

1

```
#Exercise1
#1
datind2009<-fread("datind2009.csv")
Datind2009 <- datind2009[!is.na(datind2009$age) & !is.na(datind2009$wage),]
print(cor(Datind2009$wage, Datind2009$age,method ="pearson" , use= "complete.obs"))
#The correlation between X and Y is -0.1788512
```

```
> print(cor(Datind2009$wage, Datind2009$age,method ="pearson" , use= "complete.obs"))
[1] -0.1788512
```

2、

```
#2
x<-Datind2009$age %>% as.matrix()
intercept = matrix(rep(1,20232))#there are 20232obs(lines)
X<-cbind(intercept,x)
Y<-matrix(Datind2009$wage)
Inverse<-solve(t(X)%*%X)
Beta<-Inverse%*%t(X)%*%Y
#use %*% to inner product two matrix, use "solve" to get the inverse of a matrix
print(Beta)
#Beta_hat equals to -180.1765
```

```
> print(Beta)
           [,1]
[1,] 22075.1066
[2,]  -180.1765
```

3、

```
#3
#Standard formulas
Beta[1,1]
Beta[2,1]
Y_hat<- as.matrix((Beta[2,1]*Datind2009$age-Beta[1,1]))
error<-(Y-Y_hat)
Error2<- t(error)%*%error/(length(Datind2009$age-1))
Varianceofbeta<- Error2%*%solve(t(Datind2009$age)%*%Datind2009$age)
Sd<-sqrt(Varianceofbeta)
print(Sd)
#the sd of beta age is 6.560566

#bootstrap 49
R=49
nind=length(Datind2009$age)
Return<- mat.or.vec(49, 2)
set.seed(5201314)
for (i in 1:R) {
  sample<-sample(1:nind,nind,rep=TRUE)
  data_sample<-Datind2009[sample,]
  x_boot<-data_sample$age %>% as.matrix()
  intercept = matrix(rep(1,20232))
  X_boot<-cbind(intercept,x_boot)
  Y_boot<-matrix(data_sample$wage)
  Inverse<-solve(t(X_boot)%*%X_boot)
  Beta_boot1<-Inverse%*%t(X_boot)%*%Y_boot

  Return[i,]<-Beta_boot1
}
mean_Return1<-apply(Return, 2, mean)
sd_Return1<-apply(Return, 2, sd)
#the sd of beta of age is 5.127342 and the mean beta age is -179.4399
```

```
#bootstrap499
R=499
nind=length(Datind2009$age)
Return<- mat.or.vec(499, 2)
for (i in 1:R) {
  sample<-sample(1:nind,nind,rep=TRUE)
  data_sample<-Datind2009[sample,]
  x_boot<-data_sample$age %>% as.matrix()
  intercept = matrix(rep(1,20232))
  X_boot<-cbind(intercept,x_boot)
  Y_boot<-matrix(data_sample$wage)
  Inverse<-solve(t(X_boot)%*%X_boot)
  Beta_boot2<-Inverse%*%t(X_boot)%*%Y_boot

  Return[i,]<-Beta_boot2
}
mean_Return2<-apply(Return, 2, mean)
sd_Return2<-apply(Return, 2, sd)
#the sd of beta of age is 4.970636 and the mean beta age is -180.287
#The mean of beta age in 499bootstrap is closer to the results generated from the standard formulas than that in 49 bootstrap.
#However, the sd of beta shows an opposite way, and the reason might be fixed if we increase the number from 499 to 4999.
```

```
> sd_Return2<-apply(Return, 2, sd)
> sd_Return2
[1] 273.696015   4.970636
```

Exercise2

1、

```
#Exercise2
datind05<-fread("datind2005.csv",colClasses=c(idind="character",idmen="character"))
datind06<-fread("datind2006.csv",colClasses=c(idind="character",idmen="character"))
datind07<-fread("datind2007.csv",colClasses=c(idind="character",idmen="character"))
datind08<-fread("datind2008.csv",colClasses=c(idind="character",idmen="character"))
datind09<-fread("datind2009.csv",colClasses=c(idind="character",idmen="character"))
datind10<-fread("datind2010.csv",colClasses=c(idind="character",idmen="character"))
datind11<-fread("datind2011.csv",colClasses=c(idind="character",idmen="character"))
datind12<-fread("datind2012.csv",colClasses=c(idind="character",idmen="character"))
datind13<-fread("datind2013.csv",colClasses=c(idind="character",idmen="character"))
datind14<-fread("datind2014.csv",colClasses=c(idind="character",idmen="character"))
datind15<-fread("datind2015.csv",colClasses=c(idind="character",idmen="character"))
datind16<-fread("datind2016.csv",colClasses=c(idind="character",idmen="character"))
datind17<-fread("datind2017.csv",colClasses=c(idind="character",idmen="character"))
datind18<-fread("datind2018.csv",colClasses=c(idind="character",idmen="character"))
append1<-rbind(datind05,datind06,datind07,datind08,datind09,datind10,datind11,datind12,datind13,datind14,datind15,d

#1
Append1<-append1[!is.na(append1$age)&!is.na(append1$wage),]
Append2 <- Append1 %>% mutate(ag = case_when( age <  18  ~ "0-18" ,
                             age >= 18 & age <= 25 ~ "18-25",
                             age >= 26 & age <= 30 ~ "26-30",
                             age >= 31 & age <= 35 ~ "31-35",
                             age >= 36 & age <= 40 ~ "36-40",
                             age >= 41 & age <= 45 ~ "41-45",
                             age >= 46 & age <= 50 ~ "46-50",
                             age >= 51 & age <= 55 ~ "51-55",
                             age >= 56 & age <= 60 ~ "56-60",
                             age >  60   ~ "60+"))
```

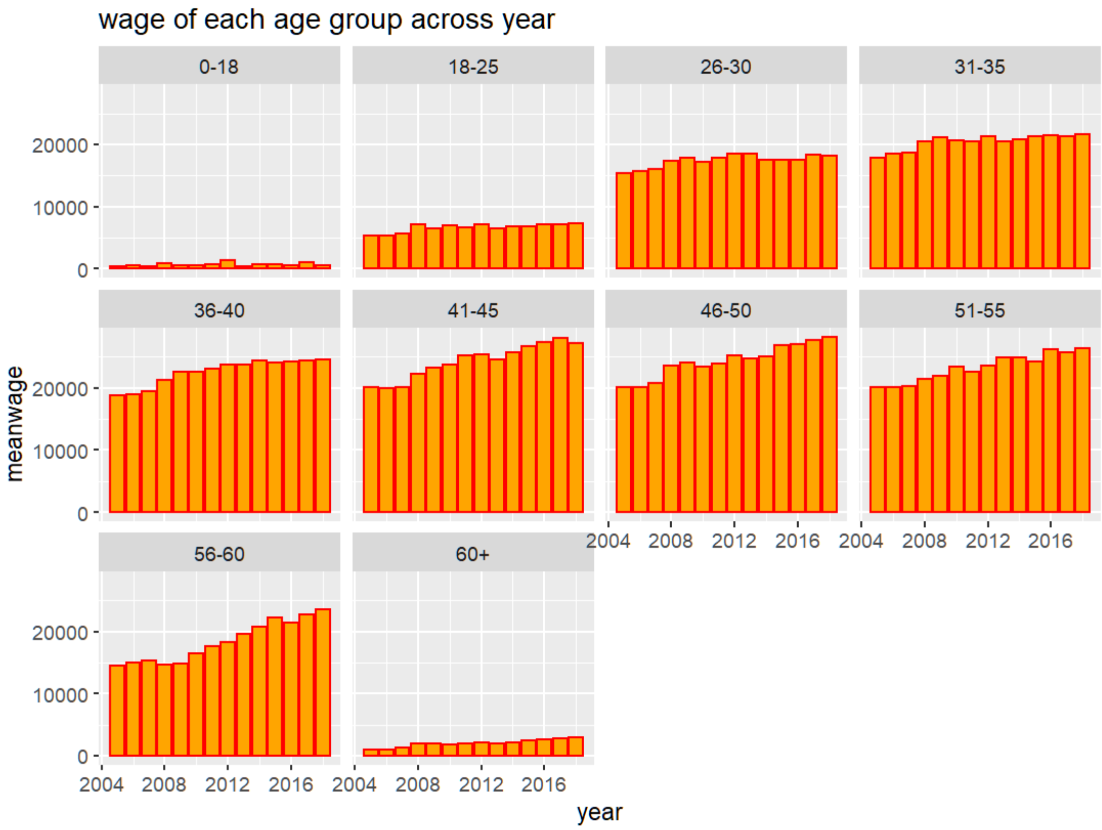
Filter

	V1	idind	idmen	year	empstat	respondent	profession	gender	age	wage	ag
1	1	1120001004058010001	1200010040580100	2005	Inactive	1		Female	31	12334	31-35
2	3	1120001006663010001	1200010066630100	2005	Employed	1	38	Male	32	50659	31-35
3	4	1120001006663010002	1200010066630100	2005	Employed	0	45	Female	28	19231	26-30
4	5	1120001008245010001	1200010082450100	2005	Retired	1		Female	90	0	60+
5	6	1120001008644010001	1200010086440100	2005	Employed	1	34	Male	37	31511	36-40
6	7	1120001008644010002	1200010086440100	2005	Employed	0	42	Female	35	24873	31-35
7	8	1120001010299010001	1200010102990100	2005	Employed	1	55	Female	41	30080	41-45
8	9	1120001010299010002	1200010102990100	2005	Inactive	0		Female	16	0	0-18
9	10	1120001011845010001	1200010118450100	2005	Employed	1	37	Male	55	43296	51-55
10	11	1120001011845010002	1200010118450100	2005	Employed	0	54	Female	55	20426	51-55
11	12	1120002001293010001	1200020012930100	2005	Employed	1	11	Male	57	0	56-60
12	13	1120002001293010002	1200020012930100	2005	Employed	0	11	Female	52	0	51-55
13	16	1120002001739010001	1200020017390100	2005	Employed	1	11	Male	51	0	51-55
14	17	1120002002642010001	1200020026420100	2005	Employed	1	11	Male	47	0	46-50
15	19	1120002004513010001	1200020045130100	2005	Employed	1	34	Female	55	49240	51-55
16	20	1120002004513010002	1200020045130100	2005	Inactive	0		Male	17	0	0-18
17	21	1120002009437010001	1200020094370100	2005	Employed	1	22	Male	41	15005	41-45
18	22	1120002009437010002	1200020094370100	2005	Employed	0	34	Female	39	35192	36-40
19	25	1120002011845010001	1200020118450100	2005	Retired	1		Male	80	0	60+
20	26	1120002012268010001	1200020122680100	2005	Employed	1	45	Male	30	22852	26-30
21	27	1120002012268010002	1200020122680100	2005	Inactive	0		Female	32	1832	31-35

"ag" is the created categorical variable.

2、

```
#2
Append3<-group_by(Append2,ag,year)
Append4<-summarise(Append3,meanwage=mean(wage,na.rm=TRUE))
plot<-ggplot(Append4,aes(x=year,y=meanwage,color = as.factor(ag)))+
             geom_bar(data = NULL, stat = "identity", width=0.9,
                      color='red',fill='orange')+
             ggtitle("wage of each age group across year")+
             theme(legend.position="top")+
             facet_wrap(~ ag)
plot
#From the plot, we can find that the wage of each age group increases as years.
#Specially, wage of "56-60" has the fastest increase and wage of age lower than 30 has the slowest increase.
#wages of "0-18" and "60+" have the lowest avg wage.
#wages of 36 to 60 are the highest and enjoy a fast increase as years.
```



#There is a trend. From the plot, we can find that the wage of each age group increases as years.

#Specially, wage of "56-60" has the fastest increase and wage of age lower than 30 has the slowest increase.

#wages of "0-18" and "60+" have the lowest avg wage.

#wages of 36 to 60 are the highest and enjoy a fast increase as years.

3、

```
#3
#X<-matrix(Append1$age,Append1$year)
X<-Append1%>%select(age,year)%>%as.matrix()
Y<-matrix(Append1$wage)
Inverse<-solve(t(X)%*%X)
Beta<-Inverse%*%t(X)%*%Y
print(Beta)
#The beta of age is -182.9581 if we add year as a variable
#if we want a time fixed effect, we need to use "plm" function
timefixedreturn <- plm(wage ~ age, data = Append1, index = "year")
summary(timefixedreturn)
#Then the beta of age equals to -186.8793
```

#Here index control for heterogeneity in variables year

```
> print(Beta)
          [,1]
age  -182.9581
year   11.2271
```

```
> summary(timefixedreturn)
Oneway (individual) effect Within Model

Call:
plm(formula = wage ~ age, data = Append1, index = "year")

Unbalanced Panel: n = 14, T = 18767-22742, N = 289769

Residuals:
     Min.   1st Qu.    Median   3rd Qu.      Max.
 -21321.1  -11464.5   -7266.0    8496.3 1738140.9

Coefficients:
     Estimate Std. Error t-value  Pr(>|t|)
age -186.8793     2.0016 -93.366 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Total Sum of Squares:    1.2219e+14
Residual Sum of Squares: 1.1862e+14
R-Squared:      0.029206
Adj. R-Squared: 0.02916
F-statistic: 8717.28 on 1 and 289754 DF, p-value: < 2.22e-16
```

Exercise3

1、

```
#Exercise3
#1
Datind2007<-datind07%>%filter(empstat!="Inactive")%>%filter(empstat!="Retired")
```

	V1	idind	idmen	year	empstat	res
1	1	1140001000124010001	1400010001240100	2007	Unemployed	
2	2	1140001000124010002	1400010001240100	2007	Employed	
3	4	1140001001167010001	1400010011670100	2007	Employed	
4	8	1140001002054010001	1400010020540100	2007	Employed	
5	9	1140001002054010002	1400010020540100	2007	Unemployed	
6	13	1140001005753010002	1400010057530100	2007	Employed	
7	14	1140001005753010003	1400010057530101	2007	Employed	
8	15	1140001007323010001	1400010073230100	2007	Employed	
9	21	1140001007351010001	1400010073510100	2007	Employed	
10	22	1140001007351010002	1400010073510100	2007	Employed	
11	23	1140001007351010003	1400010073510100	2007	Employed	
12	26	1140001010658010001	1400010106580100	2007	Employed	
13	27	1140001010658010002	1400010106580100	2007	Employed	
14	28	1140001010658010003	1400010106580100	2007	Employed	
15	32	1140001012406010001	1400010124060100	2007	Unemployed	
16	36	1140002004331010002	1400020043310100	2007	Employed	
17	41	1140002009113010001	1400020091130100	2007	Employed	
18	44	1140002012114010001	1400020121140100	2007	Employed	
19	45	[illegible]	[illegible]	2007	Employed	

This step removes all “inactive” and “retired”

2、

```
#2
#Since there are only two state of empstat:employed and unemployed
#First,we find independent variable "age", and create a dummy variable for "empstat"
age<-Datind2007$age
empstat<-Datind2007$empstat
empstat[ which(empstat == "Employed") ] = 1
empstat[ which(empstat == "Unemployed") ] = 0
empstat <- as.numeric(empstat)
likelihood<- function(coff,age,empstat)
{xbeta = coff[1] + coff[2]*age
pr  = pnorm(xbeta)
pr[pr>0.999999] = 0.999999
pr[pr<0.000001] = 0.000001
like          = empstat*log(pr) + (1-empstat)*log(1-pr)
return(-sum(like))}
```

3、

```r
#3
ntry=200
return = mat.or.vec(ntry,3)
for (i in 1:ntry){
  start    = runif(2,-5,5)
  res      = optim(start,fn=likelihood,method="BFGS",control=list(trace=6,maxit=1000),age = age, empstat = empstat)
  return[i,] = c(res$par,res$value)
}
return<-data.frame(return)
return <- format(return, scientific = F)
colnames(return) = c("beta_0", "beta_age", "likelihood_value")
return[which(return$likelihood_value == min(return$likelihood_value)),]
#We have intercept=1.045310107,Beta_age=0.006900150,likelihood=3555.890
#Since the beta_age is positive, We can say that age has a positive effect onlabor mkt participation.
```

```
> return[which(return$likelihood_value == min(return$likelihood_value)),]
         beta_0         beta_age likelihood_value
33   1.045310107       0.006900150          3555.890
```

#We have intercept=1.045310107,Beta_age=0.006900150,likelihood=3555.890

#Since the beta_age is positive, We can say that age has a positive effect onlabor mkt participation.

4、

```r
#4
Datind2007[!is.na(Datind2007$age)&!is.na(Datind2007$wage),]
Datind2007<-Datind2007%>%filter(wage!=0)
wage<-Datind2007$wage
age<-Datind2007$age
empstat<-Datind2007$empstat
empstat[ which(empstat == "Employed") ] = 1
empstat[ which(empstat == "Unemployed") ] = 0
empstat <- as.numeric(empstat)
likelihood2<- function(coff,age,wage,empstat)
{xbeta = coff[1] + coff[2]*age +coff[3]*wage
pr  = pnorm(xbeta)
pr[pr>0.999999] = 0.999999
pr[pr<0.000001] = 0.000001
like          = empstat*log(pr) + (1-empstat)*log(1-pr)
return(-sum(like))}
ntry=200
return2 = mat.or.vec(ntry,4)

for (i in 1:ntry){
  start = c(runif(1,-0.1,0.1),runif(1,-0.01,0.01),runif(1,-0.0001,0.0001))
  res   = optim(start,fn=likelihood2,method="BFGS",control=list(trace=6,maxit=1000),age = age, wage = wage, empstat = empstat)
  return2[i,] = c(res$par,res$value)
}
return2<-data.frame(return2)
return2 <- format(return2, scientific = F)
colnames(return2) = c("beta_0", "beta_age","beta_wage", "likelihood_value")
return2[which(return2$likelihood_value == min(return2$likelihood_value)),]
#We have intercept=0.0844164496,beta_age=0.0086827248,beta_wage=0.00007762600,likelihood_value=1832.81
#We find that both beta_age and beta_wage are positive, which means that age and wage have a positive effect on labor mkt participation.
```

```
> return2[which(return2$likelihood_value == min(return2$likelihood_value)),]
         beta_0       beta_age      beta_wage likelihood_value
32  0.0844164496  0.0086827248 0.00007762600         1832.814
```

#We have intercept=0.0844164496,beta_age=0.0086827248,beta_wage=0.00007762600,likelihood_value=1832.81

#We find that both beta_age and beta_wage are positive, which means that age and wage have a positive effect on labor mkt participation.

Exercise4

1、

```
#1
datind05<-fread("datind2005.csv",colClasses=c(idind="character",idmen="character"))
datind06<-fread("datind2006.csv",colClasses=c(idind="character",idmen="character"))
datind07<-fread("datind2007.csv",colClasses=c(idind="character",idmen="character"))
datind08<-fread("datind2008.csv",colClasses=c(idind="character",idmen="character"))
datind09<-fread("datind2009.csv",colClasses=c(idind="character",idmen="character"))
datind10<-fread("datind2010.csv",colClasses=c(idind="character",idmen="character"))
datind11<-fread("datind2011.csv",colClasses=c(idind="character",idmen="character"))
datind12<-fread("datind2012.csv",colClasses=c(idind="character",idmen="character"))
datind13<-fread("datind2013.csv",colClasses=c(idind="character",idmen="character"))
datind14<-fread("datind2014.csv",colClasses=c(idind="character",idmen="character"))
datind15<-fread("datind2015.csv",colClasses=c(idind="character",idmen="character"))
appendnew<-rbind(datind05,datind06,datind07,datind08,datind09,datind10,datind11,datind12,datind13,datind14,datind15)
Appendnew<-appendnew%>%filter(empstat!="Inactive")%>%filter(empstat!="Retired")
Appendnew$"2005" = as.numeric(Appendnew$year==2005)
Appendnew$"2006" = as.numeric(Appendnew$year==2006)
Appendnew$"2007" = as.numeric(Appendnew$year==2007)
Appendnew$"2008" = as.numeric(Appendnew$year==2008)
Appendnew$"2009" = as.numeric(Appendnew$year==2009)
Appendnew$"2010" = as.numeric(Appendnew$year==2010)
Appendnew$"2011" = as.numeric(Appendnew$year==2011)
Appendnew$"2012" = as.numeric(Appendnew$year==2012)
Appendnew$"2013" = as.numeric(Appendnew$year==2013)
Appendnew$"2014" = as.numeric(Appendnew$year==2014)
Appendnew$"2015" = as.numeric(Appendnew$year==2015)
age<-Appendnew$age
t06<-Appendnew$"2006" #we set 2005 as reference variable
t07<-Appendnew$"2007"
t08<-Appendnew$"2008"
t09<-Appendnew$"2009"
t10<-Appendnew$"2010"
```

```
age<-Appendnew$age
t06<-Appendnew$"2006" #we set 2005 as reference variable
t07<-Appendnew$"2007"
t08<-Appendnew$"2008"
t09<-Appendnew$"2009"
t10<-Appendnew$"2010"
t11<-Appendnew$"2011"
t12<-Appendnew$"2012"
t13<-Appendnew$"2013"
t14<-Appendnew$"2014"
t15<-Appendnew$"2015"
```

	V1	idind	idmen	year	empstat	respondent	profession	gender	age	wage	2005	2006	2007
1	3	1120001006663010001	1200010066630100	2005	Employed	1	38	Male	32	50659	1	0	
2	4	1120001006663010002	1200010066630100	2005	Employed	0	45	Female	28	19231	1	0	
3	6	1120001008644010001	1200010086440100	2005	Employed	1	34	Male	37	31511	1	0	
4	7	1120001008644010002	1200010086440100	2005	Employed	0	42	Female	35	24873	1	0	
5	8	1120001010299010001	1200010102990100	2005	Employed	1	55	Female	41	30080	1	0	
6	10	1120001011845010001	1200010118450100	2005	Employed	1	37	Male	55	43296	1	0	
7	11	1120001011845010002	1200010118450100	2005	Employed	0	54	Female	55	20426	1	0	
8	12	1120002001293010001	1200020012930100	2005	Employed	1	11	Male	57	0	1	0	
9	13	1120002001293010002	1200020012930100	2005	Employed	0	11	Female	52	0	1	0	
10	15	1120002001293010004	1200020012930100	2005	Employed	0		Male	50	NA	1	0	
11	16	1120002001739010001	1200020017390100	2005	Employed	1	11	Male	51	0	1	0	
12	17	1120002002642010001	1200020026420100	2005	Employed	1	11	Male	47	0	1	0	
13	18	1120002002642010002	1200020026420100	2005	Employed	0		Male	39	NA	1	0	
14	19	1120002004513010001	1200020045130100	2005	Employed	1	34	Female	55	49240	1	0	0
15	21	1120002009437010001	1200020094370100	2005	Employed	1	22	Male	41	15005	1	0	
16	22	1120002009437010002	1200020094370100	2005	Employed	0	34	Female	39	35192	1	0	
17	26	1120002012268010001	1200020122680100	2005	Employed	1	45	Male	30	22852	1	0	
18	29	1120014901293010001	1200149012930100	2005	Employed	1	46	Male	42	28247	1	0	
19	30	1120014901293010002	1200149012930100	2005	Employed	0	55	Female	36	21134	1	0	
20	38	1120014909940010002	1200149099400100	2005	Unemployed	0		Male	41	0	1	0	

2、

For probit

```
#probit model
empstat<-Appendnew$empstat
empstat[ which(empstat == "Employed") ] = 1
empstat[ which(empstat == "Unemployed") ] = 0
empstat <- as.numeric(empstat)
likelihood3 = function(coff,age,t06,t07,t08,t09,t10,t11,t12,t13,t14,t15,empstat)
{xbeta = coff[1] + coff[2]*age +coff[3]*t06+coff[4]*t07+coff[5]*t08+coff[6]*t09+
  coff[7]*t10+coff[8]*t11+coff[9]*t12+coff[10]*t13+coff[11]*t14+coff[12]*t15
pr   = pnorm(xbeta)
pr[pr>0.999999]  = 0.999999
pr[pr<0.000001]  = 0.000001
like             = empstat*log(pr) + (1-empstat)*log(1-pr)
return(-sum(like))}

ntry = 100
return3 = mat.or.vec(ntry,13)

for (i in 1:ntry){
  start = runif(12, -5, 5)
  res = optim(start,fn = likelihood3,method = "BFGS",
              control = list(trace=6,maxit=1000),
              age = age,t06=t06,t07=t07,t08=t08,t09=t09,t10=t10,t11=t11,t12=t12,
              t13=t13,t14=t14,t15=t15,empstat = empstat)
  return3[i,] = c(res$par, res$value)}

return3 = data.frame(return3)
colnames(return3) = c("beta_0", "beta_age","beta_5","beta_6","beta_7","beta_8","beta_9","beta_10","beta_11","beta_12",
                      "beta_13","beta_14","likelihood")
return3[which(return3$likelihood == min(return3$likelihood)),]
#After adding fixed time effect, beta age is 0.01231345 which is positive.
#which means that age have a positive effect on labor mkt participation.
```

```
> return3[which(return3$likelihood == min(return3$likelihood)),]
      beta_0   beta_age     beta_5     beta_6    beta_7     beta_8     beta_9    beta_10     beta_11     beta_12
86 0.7498015 0.01231345 0.01683352 0.0799461 0.1099676 0.02610837 0.02162842 0.05490909 0.009367533 -0.04021522
       beta_13     beta_14 likelihood
86 -0.03273255 -0.05343145   42243.66
```

For logit

```
#Linear
likelihood5 = function(coff,age,t06,t07,t08,t09,t10,t11,t12,t13,t14,t15,empstat)
{xbeta = coff[1] + coff[2]*age +coff[3]*t06+coff[4]*t07+coff[5]*t08+coff[6]*t09+
  coff[7]*t10+coff[8]*t11+coff[9]*t12+coff[10]*t13+coff[11]*t14+coff[12]*t15
pr   = xbeta
pr[pr>0.999999]  = 0.999999
pr[pr<0.000001]  = 0.000001
like             = empstat*log(pr) + (1-empstat)*log(1-pr)
return(-sum(like))}

ntry = 100
return5 = mat.or.vec(ntry,13)

for (i in 1:ntry){
  start = runif(12, -5, 5)
  res = optim(start,fn = likelihood5,method = "BFGS",
              control = list(trace=6,maxit=1000),
              age = age,t06=t06,t07=t07,t08=t08,t09=t09,t10=t10,t11=t11,t12=t12,
              t13=t13,t14=t14,t15=t15,empstat = empstat)
  return5[i,] = c(res$par, res$value)}

return5 = data.frame(return5)
colnames(return5) = c("beta_0", "beta_age","beta_5","beta_6","beta_7","beta_8","beta_9","beta_10","beta_11","beta_12",
                         "beta_13","beta_14","likelihood")
return5[which(return5$likelihood == min(return5$likelihood)),]
#After adding fixed time effect, beta age is -0.04029524 which is negative.
#which means that age have a negative effect on labor mkt participation.
#If we increases ntry to 1000, the beta_age might change.
```

```
> return5[which(return5$likelihood == min(return5$likelihood)),]
     beta_0    beta_age  beta_5   beta_6   beta_7     beta_8    beta_9  beta_10  beta_11    beta_12  beta_13    beta_14
85 5.954075 -0.04029524 4.45932 2.524891 1.931448 -0.0331935 -3.069214 3.223872 2.456985 0.06812651 1.170139 -0.2420417
   likelihood
85   180372.2
```

#After adding fixed time effect, beta age is -0.04029524 which is negative.

#which means that age have a negative effect on labor mkt participation.

#If we increases ntry to 1000, the beta_age might change.

3、

```
#3
#The parameters generated by porbit and logit can be interpreted as whether there is positive or negative relationship
#between age and labor mkt participation.
#The parameter generated by the OLS model can be interpreted as certain degree of change of labor mkt participation
#as one unit change of age.
```

Exercise5

1 、

```
#Exercise5
#1
intercept = matrix(rep(1,128636))#there are 128636bs(lines)
X<-cbind(intercept,age,t06,t07,t08,t09,t10,t11,t12,t13,t14,t15)
install.packages("margins")
library(margins)
mar1<-glm(empstat~age+t06+t07+t08+t09+t10+t11+t12+t13+t14+t15,family=binomial(link=probit),data=Appendnew)
summary(mar1)
marg1probit<-margins(mar1)
mar2<-glm(empstat~age+t06+t07+t08+t09+t10+t11+t12+t13+t14+t15,family=binomial(link=logit),data=Appendnew)
summary(mar2)
marg1logit<-margins(mar2)

#2
set.seed(12345)
bootmarg <- mat.or.vec(30, 12)
```