```
# Assignment 2
# Exercise 1
# Q1 correlation is 0.09337049
X = datind2009[,9]
Y = datind2009[,10]
X = as.vector(X)
Y = as.vector(Y)
index1 = is.na(Y)
Y[index1] = 0
corr1 = sum((X-mean(X))*(Y-mean(Y)))
corr2 = (sum((X-mean(X))^2)*sum((Y-mean(Y))^2))^(1/2)
corr = corr1/corr2
# Q2 Beta is 218.0255
X = as.matrix(X)
Y = as.matrix(Y)
beta = solve(t(X)%*%X)%*%t(X)%*%Y
# Q3 Standard deviation is 4.874906e-06
e = Y-218.0255*X
Varbeta = solve(t(X)%*%X)%*%t(X)%*%(e%*%t(e))%*%X%*%solve(t(X)%*%X)
Varbeta = as.numeric(Varbeta)
stbeta = Varbeta^(1/2)
# Bootstrap 49 Replications
YX = cbind(Y,X)
YX = as.data.frame(YX)
N = 49
nind = length(X)
nvar = 2
outs = mat.or.vec(N,nvar)
set.seed(123)
```

```r
for (i in 1:N)
{
  samp     = sample(1:nind,nind,rep=TRUE)
  dat_samp = YX[samp,]
  reg     = lm(V1~V2,data = dat_samp)
  outs[i,] = reg$coefficients
}
sdest = apply(outs, 2, sd)
# Bootstrap 499 Replications
N1 = 499
nind1 = length(X)
nvar1 = 2
outs1 = mat.or.vec(N1,nvar1)
set.seed(123)
for (i in 1:N1)
{
  samp     = sample(1:nind1,nind1,rep=TRUE)
  dat_samp = YX[samp,]
  reg1     = lm(V1~V2,data = dat_samp)
  outs1[i,] = reg1$coefficients
}
sdest1 = apply(outs1, 2, sd)
est = cbind(summary(reg)$coefficients[,2],sdest,sdest1)
# Bootstrap with 499 replications has beta 3.576229, closer to estimates from lm which is 4.70695

# Exercise 2
# Q1
```

```
ind0518 =
bind_rows(datind2005,datind2006,datind2007,datind2008,datind2009,datind2010,datind2011,datind20
12,datind2013,datind2014,datind2015,datind2016,datind2017,datind2018)

index2 = is.na(ind0518)

ind0518 = ind0518[,c(4,9,10)]

ind0518 = drop_na(ind0518)

ind0518 = ind0518[ind0518$wage != 0,]

ind0518 = ind0518[ind0518$age >= 18,]

ind0518$ag = cut(ind0518$age, c(17,25,30,35,40,45,50,55,60,200))

age1 = ind0518 %>% filter(ag == "(17,25]") %>% group_by(year) %>% summarize(mean_wage =
mean(wage))

age2 = ind0518 %>% filter(ag == "(25,30]") %>% group_by(year) %>% summarize(mean_wage =
mean(wage))

age3 = ind0518 %>% filter(ag == "(30,35]") %>% group_by(year) %>% summarize(mean_wage =
mean(wage))

age4 = ind0518 %>% filter(ag == "(35,40]") %>% group_by(year) %>% summarize(mean_wage =
mean(wage))

age5 = ind0518 %>% filter(ag == "(40,45]") %>% group_by(year) %>% summarize(mean_wage =
mean(wage))

age6 = ind0518 %>% filter(ag == "(45,50]") %>% group_by(year) %>% summarize(mean_wage =
mean(wage))

age7 = ind0518 %>% filter(ag == "(50,55]") %>% group_by(year) %>% summarize(mean_wage =
mean(wage))

age8 = ind0518 %>% filter(ag == "(55,60]") %>% group_by(year) %>% summarize(mean_wage =
mean(wage))

age9 = ind0518 %>% filter(ag == "(60,200]") %>% group_by(year) %>% summarize(mean_wage =
mean(wage))

ggplot(age1, aes(x=year, y=mean_wage))+geom_line()+ylab("Average income")

ggplot(age2, aes(x=year, y=mean_wage))+geom_line()+ylab("Average income")

ggplot(age3, aes(x=year, y=mean_wage))+geom_line()+ylab("Average income")

ggplot(NULL,aes(x=year,y=mean_wage))+geom_line(data=age1,col="red")+geom_line(data=age2,col="b
lue")+geom_line(data=age3,col="green")

ggplot(age4, aes(x=year, y=mean_wage))+geom_line()+ylab("Average income")
```

```
ggplot(age5, aes(x=year, y=mean_wage))+geom_line()+ylab("Average income")

ggplot(age6, aes(x=year, y=mean_wage))+geom_line()+ylab("Average income")

ggplot(age7, aes(x=year, y=mean_wage))+geom_line()+ylab("Average income")

ggplot(age8, aes(x=year, y=mean_wage))+geom_line()+ylab("Average income")

ggplot(age9, aes(x=year, y=mean_wage))+geom_line()+ylab("Average income")

ggplot(NULL,aes(x=year,y=mean_wage))+geom_line(data=age1,col="red")+geom_line(data=age2,col="blue")+geom_line(data=age3,col="green")+geom_line(data=age4,col="yellow")+geom_line(data=age5,col="brown")+geom_line(data=age6,col="black")+geom_line(data=age7,col="darkorange")+geom_line(data=age8,col="purple")+geom_line(data=age9,col="darkgray")

# Q2 After control for time fixed effect, age and wage are more positively correlated. The coefficient is 297.9, it is also statistically significant.

ind0518$year = as.factor(ind0518$year)

lmtifix = lm(wage ~ age + year-1, data = ind0518)


# Exercise 3

# Q1

E3Q1 = datind2007[datind2007$empstat != "Inactive",]

E3Q1 = E3Q1[E3Q1$empstat != "Retired",]

E3Q1$empstat = ifelse(E3Q1$empstat == "Employed",1,0)

# Q2 & Q3

x1 = E3Q1$age

y1 = E3Q1$empstat

set.seed(123)

beta = runif(2)

flike = function(beta,x,y)

{

  xbeta = beta[1] + x*beta[2]

  pr            = pnorm(xbeta)

  pr[pr>0.999999] = 0.999999

  pr[pr<0.000001] = 0.000001
```

```
  like          = y*log(pr) + (1-y)*log(1-pr)
  return(-sum(like))
}
flike(beta,x1,y1) # log likelihood is 11277.02
N2 = 100
out1 = mat.or.vec(N2,2)
for (i0 in 1:N2)
{
  begin    = runif(2,-10,10)
  res      = optim(begin,fn=flike,method="BFGS",control=list(trace=6,maxit=1000),x=x1,y=y1)
  out1[i0,] = res$par
}
start = runif(2)
res  =
optim(start,fn=flike,method="BFGS",control=list(trace=6,REPORT=1,maxit=1000),x=x1,y=y1,hessian=TR
UE)
res$par # beta is estimated to be 0.6676402


# Exercise 4
# Q1
ind0515 =
bind_rows(datind2005,datind2006,datind2007,datind2008,datind2009,datind2010,datind2011,datind20
12,datind2013,datind2014,datind2015)
ind0515 = ind0515[ind0515$empstat != "Inactive",]
ind0515 = ind0515[ind0515$empstat != "Retired",]
ind0515$empstat = ifelse(ind0515$empstat == "Employed",1,0)
# Q2
# Probit Model
x2 = ind0515$age
y2 = ind0515$empstat
```

```
z2 = ind0515$year - 1
beta2 = runif(3)
flike2 = function(beta,x,z,y)
{
  xbeta = beta[1] + beta[2]*x + beta[3]*z
  pr         = pnorm(xbeta)
  pr[pr>0.999999] = 0.999999
  pr[pr<0.000001] = 0.000001
  like       = y*log(pr) + (1-y)*log(1-pr)
  return(-sum(like))
}
flike2(beta2,x2,z2,y2) # 183608.3
N3 = 100
out2 = mat.or.vec(N3,3)
for (i0 in 1:N3)
{
  begin   = runif(3,-10,10)
  res1    = optim(begin,fn=flike2,method="BFGS",control=list(trace=6,maxit=1000),x=x2,z=z2,y=y2)
  out2[i0,] = res1$par
}
start = runif(3)
res1  =
optim(start,fn=flike2,method="BFGS",control=list(trace=6,REPORT=1,maxit=1000),x=x2,z=z2,y=y2,hessia
n=TRUE)
res1$par # beta = 0.2991228
# Logit Model
x2 = ind0515$age
y2 = ind0515$empstat
beta3 = runif(3)
```

```
flike3 = function(beta,x,z,y)

{

  xbeta = beta[1] + beta[2]*x + beta[3]*z

  pr             = exp(beta)/(1+exp(beta))

  pr[pr>0.999999] = 0.999999

  pr[pr<0.000001] = 0.000001

  like          = y*log(pr) + (1-y)*log(1-pr)

  return(-sum(like))

}

flike3(beta3,x2,z2,y2) # log likelihood is 65758.75

N3 = 100

out2 = mat.or.vec(N3,3)

for (i0 in 1:N3)

{

  begin    = runif(2,-10,10)

  res2      = optim(begin,fn=flike3,method="BFGS",control=list(trace=6,maxit=1000),x=x2,z=z2,y=y2)

  out1[i0,] = res$par

}

start = runif(3)

res2  =
optim(start,fn=flike3,method="BFGS",control=list(trace=6,REPORT=1,maxit=1000),x=x2,z=z2,y=y2,hessia
n=TRUE)

res2$par

# linear model

EX4LM = lm(y2~x2+z2)


# Exercise 5
```