

```

# Assignment 2

# Exercise 1

# Q1 correlation is 0.06086454

X0 = datind2009[,9]
Y = datind2009[,10]

intercept = rep.int(1,length(X))

X = cbind(intercept,X0)

data.frame(X)

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 70.71433

X = as.matrix(X)
Y = as.matrix(Y)

beta = solve(t(X)%*%X)%*%t(X)%*%Y

# Q3 Standard deviation is 5.020993

e = Y-70.71433*X[,2]-7848.18701

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(Y)

nvar = 2

outs = mat.or.vec(N,nvar)

```

```

set.seed(123)
for (i in 1:N)
{
  samp  = sample(1:nind,nind,rep=TRUE)
  dat_samp = YX[samp,]
  reg  = lm(V1~X0,data = dat_samp)
  outs[i,] = reg$coefficients
}

sdest = apply(outs, 2, sd) # Sd is estimated to be 3.043132

# 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~X0,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 3.547175, 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)

lmtifx = 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

```

```

t1 = ifelse(ind0515$year == 2006, 1 , 0)
t2 = ifelse(ind0515$year == 2007, 1 , 0)
t3 = ifelse(ind0515$year == 2008, 1 , 0)
t4 = ifelse(ind0515$year == 2009, 1 , 0)
t5 = ifelse(ind0515$year == 2010, 1 , 0)
t6 = ifelse(ind0515$year == 2011, 1 , 0)
t7 = ifelse(ind0515$year == 2012, 1 , 0)
t8 = ifelse(ind0515$year == 2013, 1 , 0)
t9 = ifelse(ind0515$year == 2014, 1 , 0)
t10 = ifelse(ind0515$year == 2015, 1 , 0)

beta2 = runif(12)

flike2 = function(beta,x,t1,t2,t3,t4,t5,t6,t7,t8,t9,t10,y)
{
  xbeta = beta[1] + beta[2]*x + beta[3]*t1 + beta[4]*t2 + beta[5]*t3 + beta[6]*t4 + beta[7]*t5 +
beta[8]*t6 + beta[9]*t7 + beta[10]*t8 + beta[11]*t9 + beta[12]*t10

  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,t1,t2,t3,t4,t5,t6,t7,t8,t9,t10,y2) # 183608.3

N3 = 100

out2 = mat.or.vec(N3,12)

for (i0 in 1:N3)
{
  begin  = runif(12,-10,10)

  res1   =
optim(begin,fn=flike2,method="BFGS",control=list(trace=6,maxit=1000),x=x2,t1=t1,t2=t2,t3=t3,t4=t4,t5
=t5,t6=t6,t7=t7,t8=t8,t9=t9,t10=t10,y=y2)

```

```

    out2[i0,] = res1$par
}

start = runif(12)

res1 =
optim(start,fn=flike2,method="BFGS",control=list(trace=6,REPORT=1,maxit=1000),x=x2,t1=t1,t2=t2,t3=t
3,t4=t4,t5=t5,t6=t6,t7=t7,t8=t8,t9=t9,t10=t10,y=y2,hessian=TRUE)

res1$par # beta = 0.01231541

# Logit Model

x2 = ind0515$age
y2 = ind0515$sempstat

beta3 = runif(3)

flike3 = function(beta,x,z,y)
{
    pr      = exp(x)/(1+exp(x))
    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 183608.3

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,hessian=TRUE)

res2$par # beta is 0.7037228

# linear model

EX4LM = lm(y2~x2+factor(year),data = ind0515) # Beta is 0.002330


# Exercise 5

# Q1

# Probit model

probitb =
c(0.75073029,0.01231541,0.01509295,0.07880319,0.10827731,0.02428734,0.02084246,0.05339361,0.01173682,-0.04104253,-0.03575343,-0.05625736)

predictedprobit =
probitb[1]+probitb[2]*x2+probitb[3]*t1+probitb[4]*t2+probitb[5]*t3+probitb[6]*t4+probitb[7]*t5+probitb[8]*t6+probitb[9]*t7+probitb[10]*t8+probitb[11]*t9+probitb[12]*t10

pdf1 = pnorm(predictedprobit)

marginalprobit = 0.01231541*sum(pdf1)/length(pdf1)

marginalprobit # marginal effect is 0.01104241

# Logit Model

logitb = c(0.5154099, 0.7037228)

predictedlogit = logitb[1]+logitb[2]*x2

pdf2 = pnorm(predictedlogit)

marginalprobit = logitb[2]*sum(pdf2)/length(pdf2)

marginalprobit # marginal effect is 0.7037228

```