```
getwd()
setwd("C:/Users/86185/Dropbox/Papers/Progress")
datstu = read.csv("datstu.csv")
datsss = read.csv("datsss.csv")
datjss = read.csv("datjss.csv")
summary(datstu)
summary(datsss)
summary(datjss)
library(AER)
library(tidyr)
library(dplyr)
library(tidyverse)
#Exercise 1------
#1 Number of students
map(datstu,~sum(is.na(.)))
q1.1 = nrow(datstu)
cat('1.1 The number of students is:', q1.1, '\n')
1.1 The number of students is: 340823
#2 Number of Schools
number_of_schools <- unique(datsss$schoolcode)</pre>
number_of_schools <- data.frame(number_of_schools)</pre>
q1.2 = nrow(number_of_schools)
cat('1.2 The number of school is:',q1.2,'\n')
1.2 The number of school is: 898
#3 Number of programs
program = select(datstu, choicepgm1, choicepgm2, choicepgm3, choicepgm4, choicepgm5,
choicepgm6)
sum prgm = gather(program, choice program, program name, choicepgm1, choicepgm2,
choicepgm3, choicepgm4, choicepgm5, choicepgm6)
sum_prgm_unique = unique(sum_prgm)
cat('1.3 The number of program is:', nrow(sum_prgm_unique), '\n')
1.3 The number of program is: 174
#4 Number of choices (school,program)
school = datstu %>%
 select(schoolcode1,
                        schoolcode2.
                                        schoolcode3.
                                                        schoolcode4.
                                                                        schoolcode5,
schoolcode6) %>%
  gather(school, schoolcode, schoolcode1, schoolcode2, schoolcode3, schoolcode4,
schoolcode5, schoolcode6)
choices = cbind(school[,2], sum_prgm[,2])
choices_unique = unique(choices)
cat('1.4 The number of choices is:', nrow(choices_unique), '\n')
1.4 The number of choices is: 3086
```

```
#5 Missing test score
missing_test_score = sum(is.na(datstu$score) == 'TRUE', na.rm = TRUE)
cat('1.5 The number of missed test score is:', missing_test_score, '\n')
1.5 The number of missed test score is: 179887
#6 Apply to the same school (different programs)
f1 <- function(x) {length(x[!is.na(x)]) - length(unique(x[!is.na(x)]))}
same_school = apply(datstu[, 5:10], MARGIN = 1, FUN = f1)
length(same_school[same_school != 0])
cat('1.6 The number of students apply to the same school but different programs:',
length(same_school[same_school != 0]), '\n')
1.6 The number of students apply to the same school but different programs: 120071
#7 Apply to less than 6 choices
less_than_6 = sum(is.na(datstu$schoolcode6)=='TRUE', na.rm = TRUE)
cat('1.7 The number of student apply to less than 6 choices is:', less than 6, '\n')
1.7 The number of student apply to less than 6 choices is: 17088
#Exercise 2-----
datstu$admitted_by_schoolcode=ifelse(datstu$rankplace==1, datstu$schoolcode1,
                             ifelse(datstu$rankplace==2, datstu$schoolcode2,
                             ifelse(datstu$rankplace==3, datstu$schoolcode3,
                             ifelse(datstu$rankplace==4, datstu$schoolcode4,
                             ifelse(datstu$rankplace==5, datstu$schoolcode5,
                             ifelse(datstu$rankplace==6, datstu$schoolcode6, NA)))))
datstu$admitted=ifelse(datstu$rankplace==1, datstu$choicepgm1,
                 ifelse(datstu$rankplace==2, datstu$choicepgm2,
                 ifelse(datstu$rankplace==3, datstu$choicepgm3,
                 ifelse(datstu$rankplace==4, datstu$choicepgm4,
                 ifelse(datstu$rankplace==5, datstu$choicepgm5,
                 ifelse(datstu$rankplace==6, datstu$choicepgm6, NA))))))
data_raw=datstu %>%
  group_by(admitted) %>%
  summarise(schoolcode=admitted_by_schoolcode,
                                                                 minscore=min(score),
average=mean(score), number=n())
dataraw=data_raw %>%
  rename(school_program = admitted)
dataraw=unique(dataraw)
datsss$X=NULL
datsss=unique(datsss)
data_q2<-merge(x=dataraw,y=datsss,by="schoolcode",all.x=TRUE)
data_q2=na.omit(data_q2)
names(data_q2)[names(data_q2) == "minscore"] <- "Cutoff"</pre>
names(data_q2)[names(data_q2) == "average"] <- "Quality"</pre>
names(data_q2)[names(data_q2) == "number"] <- "Size"
```

•	schoolcode <sup>‡</sup>	school_program	Cutoff <sup>‡</sup>	Quality <sup>‡</sup>	Size <sup>‡</sup>	schoolname	sssdistrict	ssslong	ssslat
1	10101	Home Economics	190	280.0537	14994	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN	Accra Metropolitan	-0.19711526	5.607396
3	10101	Agriculture	188	274.7164	13168	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN	Accra Metropolitan	-0.19711526	5.607396
5	10101	General Science	158	334.1679	15125	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN	Accra Metropolitan	-0.19711526	5.607396
7	10101	Business	194	297.6537	30304	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN	Accra Metropolitan	-0.19711526	5.607396
9	10101	Visual Arts	173	295.9655	9989	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN	Accra Metropolitan	-0.19711526	5.607396
11	10101	General Arts	194	298.1614	46517	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN	Accra Metropolitan	-0.19711526	5.607396
13	10102	General Arts	194	298.1614	46517	ST. MARY'S SENIOR HIGH. SCHOOL, KORLE GONNO	Accra Metropolitan	-0.19711526	5.607396
15	10102	Visual Arts	173	295.9655	9989	ST. MARY'S SENIOR HIGH. SCHOOL, KORLE GONNO	Accra Metropolitan	-0.19711526	5.607396
17	10102	General Science	158	334.1679	15125	ST. MARY'S SENIOR HIGH. SCHOOL, KORLE GONNO	Accra Metropolitan	-0.19711526	5.607396
19	10102	Home Economics	190	280.0537	14994	ST. MARY'S SENIOR HIGH. SCHOOL, KORLE GONNO	Accra Metropolitan	-0.19711526	5.607396
22	10103	Agriculture	188	274.7164	13168	WESLEY GRAMMAR SCHOOL, DANSOMAN	Accra Metropolitan	-0.19711526	5.607396
24	10103	General Arts	194	298.1614	46517	WESLEY GRAMMAR SCHOOL, DANSOMAN	Accra Metropolitan	-0.19711526	5.607396
26	10103	Home Economics	190	280.0537	14994	WESLEY GRAMMAR SCHOOL, DANSOMAN	Accra Metropolitan	-0.19711526	5.607396
28	10103	Business	194	297.6537	30304	WESLEY GRAMMAR SCHOOL, DANSOMAN	Accra Metropolitan	-0.19711526	5.607396
30	10103	General Science	158	334.1679	15125	WESLEY GRAMMAR SCHOOL, DANSOMAN	Accra Metropolitan	-0.19711526	5.607396
32	10103	Visual Arts	173	295.9655	9989	WESLEY GRAMMAR SCHOOL, DANSOMAN	Accra Metropolitan	-0.19711526	5.607396
33	10104	General Arts	194	298.1614	46517	HOLY TRINITY CATHEDRAL SENIOR HIGH SCH, ACCRA	Accra Metropolitan	-0.19711526	5.607396
35	10104	General Science	158	334.1679	15125	HOLY TRINITY CATHEDRAL SENIOR HIGH SCH, ACCRA	Accra Metropolitan	-0.19711526	5.607396
37	10104	Visual Arts	173	295.9655	9989	HOLY TRINITY CATHEDRAL SENIOR HIGH SCH, ACCRA	Accra Metropolitan	-0.19711526	5.607396
39	10104	Home Economics	190	280.0537	14994	HOLY TRINITY CATHEDRAL SENIOR HIGH SCH, ACCRA	Accra Metropolitan	-0.19711526	5.607396

## #Exercise 3------

datjss=select(datjss, -X)

datstu\_jss=left\_join(datstu, datjss, by="jssdistrict")

 $datstu\_jss\_sss=left\_join(datstu\_jss, datsss, by=c("admitted\_by\_schoolcode"="schoolcode"))$ 

datstu\_jss\_sss\$distance = sqrt(

 $sqrt((69.172*(datstu\_jss\_sss\$ssslong-$ 

1		765	16	0	50112	Socot	50000	50000	50702	5090	Home Economics	General Arts	Visual Arts	Visual Arts	Home Sconomics	General Arts	Bosantue/Etx/ma/Guanuama (Curtanase)	. 16	101 10	+1.86275169	6.55532	144	AM	764	701	194
2	2	7/4	17	0	72102	70602	70107	70106	70605	7060	General Arts	Dusiness	General Arts	General Arts	Home Sconomics	General Arts	He Municipal	. 76	100	052514224	6,71760	100	7/4	764	752	.68
3	3	765	79	0	50700	50708	50115	50706	51608	5070	Business	Home Economics	Buinesi	Home Economics	Home Sconomics	Business	Svatire (Manuscrieng)	76	105 105	-1.84142010	6.80677	600	844	AA.	761	765
4	4	764	25		90501	50405	90101	9090401	90102	9050	Visua Arts	Ceneral Arts	Agriculture	Mater Vehicle Mech.	Agriculture	Ceneral Arts	Kessena/Nankani (Navrange)	- 14	14 74	-121744096	10.90942	All .	69	/44	159	368
5	5	704	15	0	57802	51701	50206	50007	51602	5000	Home (conomics)	General Arts	Home Economics	General Arts	General Arts	Home Economics	Abvirra Mponus (Nymatin)	. 16	101 101	-217718053	6,54950	AM.	NA.	764	108	768
6	6	NI	15	0	10100	50105	\$1701	50000	50601	5160	General Arts	General Arts	General Arts	General Arts	Home Boonamics	Home Bonomics	Kumasi Metro	. 16	W. 761	-1,59718716	6.68206	A44	761	765	191	.501
7	7	AR	22		80501	80401	80502	80400	80501	8090	Ceneral Arts	Ceneral Arts	General Arts	General Arts	General Arts	General Arts	Nanumba North (Simbilia)		50 704	-0.14176416	8,81677	66	666	NA	769	164
	0	NA	19	7	40001	40401	40402	40002	40202	4000	General Arts	General Arts	General Arts	Agriculture	Agriculture	Agriculture	Jomoro (Haif Assin)	- 4	101 101	-2.60022027	5,00950	NA.	NA.	NA.	59	164
9	9	NI	19	1	21303	21308	21201	21201	20208	2010	Business	Business	General Science	General Science	General Arts	General Arts	East, Akim (Klbi)	. 14	VI IVI	-0.45434421	6,17955	100	NA	Alt	N	765
10	10	701	16	0	80101	90401	50505	50901	50501	5050	General Arts	General Arts	General Arts	General Arts	General Arts	General Arts	Eura/Seigedumase (Eura)	. 16	sol not	+1.56796534	7.46265	604	504	100	59	765
11	11	704	17	0	51902	50601	50500	50602	50600	5090	Home Sconomics	Visial Arts	Home Economics	Home Sconomics	General Arts	General Arts	Sergere West (ntempong)	. 16	NA FOR	+1.10007694	7.19956	AM.	508	708	rive.	758
12	12	701	17	1	100201	90606	80107	90501	90001	9060	General Science	General Science	General Science	General Science	General Science	Agriculture	Kessera/Nankani (Nevrongo)	N	94 /94	-121744096	10.80942	NA.	AR	.04	791	761
13.	13	701	16		30608	30603	30904	30904	30602	3090	Business	Service	General Arts	Business	Business	General Arts	Agona Swedru	. 14	201 201	-075534253	5.61735	100	Alt	766	700	345
14	14	NA	76	0	80101	80102	90104	80100	80401	\$100	Dusiness	Dusiness	Home Sconomics	Home Economics	Apriculture	Agriculture	Teron Kunbungu (Teron)		101 10	-1.10971997	9.52724	704	NA	704	768	100
15	15	NA	21	- 1	90801	90602	80106	90601	10108	1011	Technical	Secretar	Technical	Agriculture	Technical	Vous Arts	Acca Metropolitan		98 NR	-0.19711526	5.60739	NA.	NA.	AA	398	765
16	16	764	17		40903	40904	40901	41102	41101	4000	Ceneral Arts	Business	Business	Business	Agriculture	Ceneral Science	Mpener-stassa East (Debesse)	176	54 554	-1.89758557	5.33079	NA.	ALL	All	768	.50
17	17	769	17	- 1	80102	80101	80106	80105	50901	5090	General Arts	General Arts	Agriculture	Dusiness	General Arts	General Arts	Cura/Selgedumate (Cura)	176	NA 704	-1.36796534	7,46363	784	NA	766	768	188
10	18	TNR.	18	- 1	10401	70505	10504	20004	20801	2110	Seneral Arts	General Arts	General Arts	General Arts	Agriculture	Technical	Ga West (Amasaman)	. 16	VE 200	-0.39751053	5.66463	NA:	NA.	- 744	586	361
19	39	704	19		60501	60502	60504	60001	40902	4090	Agriculture	Ceneral Arts	Agriculture	Apriculture	Apriculture	Agriculture	Wassa Ament (Asankragus)	. 16	101 101	~2.50201793	5.72551	NA.	A14.	764	.768	188
20	20	768	10	- 1	100102	60701	90204	60701	60106	6060	General Arts	Home Sconomics	General Arts	General Arts	Home Conomics	Home Conomics	Sole	- 66	58 758	-220007523	0.62969	AN .	204	764	258	386

## #Exercise 4-----

number\_of\_choices <- unique(choices\_unique)</pre>

choices\_unique <- [!(is.na(number\_of\_choices\$schoolcode)),]</pre>

nrow(number\_of\_choices)

new\_number\_of\_choices <- data.frame(number\_of\_choices)</pre>

colnames(new\_number\_of\_choices)[colnames(new\_number\_of\_choices) == "X1"] <-

"schoolcode"

colnames(new\_number\_of\_choices)[colnames(new\_number\_of\_choices) == "X2"] <"choicepgm"</pre>

datsss2<-datsss

datsss2<-datsss2[!(is.na(datsss2\$ssslong)),]

datsss2<-datsss2[,-1]

datsss2<-as.data.frame(unique(datsss2))

datanew<-merge(x=new\_number\_of\_choices,y=datsss2,by="schoolcode",all.x = TRUE, all.y = FALSE)

```
datstunew<-datstu
datstunew<-datstunew[!(is.na(datstunew$rankplace)),]
for (i in 1:nrow(datstunew)){
  if (datstunew$rankplace[i]==1){
    datstunew$schoolcode[i]<-datstunew$schoolcode1[i]
    datstunew$choicepgm[i]<-datstunew$choicepgm1[i]
  }
  if (datstunew$rankplace[i]==2){
    datstunew$schoolcode[i]<-datstunew$schoolcode2[i]
    datstunew$choicepgm[i]<-datstunew$choicepgm2[i]
  }
  if (datstunew$rankplace[i]==3){
    datstunew$schoolcode[i]<-datstunew$schoolcode3[i]
    datstunew$choicepgm[i]<-datstunew$choicepgm3[i]
  }
  if (datstunew$rankplace[i]==4){
    datstunew$schoolcode[i]<-datstunew$schoolcode4[i]
    datstunew$choicepgm[i]<-datstunew$choicepgm4[i]
  }
  if (datstunew$rankplace[i]==5){
    datstunew$schoolcode[i]<-datstunew$schoolcode5[i]
    datstunew$choicepgm[i]<-datstunew$choicepgm5[i]
  }
  if (datstunew$rankplace[i]==6){
    datstunew$schoolcode[i]<-datstunew$schoolcode6[i]
    datstunew$choicepgm[i]<-datstunew$choicepgm6[i]
  }
datstunew<-datstunew[!(datstunew$rankplace == 99),]
data_final<-datstunew %>%
  group_by(schoolcode,choicepgm) %>%
  summarise(cutoff=min(score),quality = mean(score),size = n())
data_final<-merge(x=datanew,y=data_final,by= c("schoolcode", "choicepgm"))
datanew2<-merge(x=datstunew,y=data_final,by= c("schoolcode", "choicepgm"))
datanew2<-merge(x=datanew2,y=datjss,by="jssdistrict",all.x = TRUE, all.y = FALSE)
colnames(datanew2)[colnames(datanew2) == "point_x"] <- "jsslong"</pre>
colnames(datanew2)[colnames(datanew2) == "point_y"] <- "jsslat"
datanew2$distance <- 0
for (i in 1:nrow(datanew2)){
```

```
datanew2$distance[i]<-sqrt((69.172
                                                           (datanew2$ssslong[i]-
datanew2$jsslong[i])*cos(datanew2$jsslat[i]/57.3))^2 + (69.172 * (datanew2$ssslat[i] -
datanew2$jsslat[i]))^2)
datanew2<-datanew2[!(is.na(datanew2$score)),]
datanew2<-datanew2[!(is.na(datanew2$distance)),]
datanew2 %>%
 group_by(rankplace) %>%
 summarise(cutoff=min(score),quality = mean(score),distance=mean(distance))
# A tibble: 6 x 4
   rankplace cutoff quality distance
                               <db7>
                                            <db7>
         <int>
                   <1nt>
1
               1
                      165
                                314.
                                             35.2
2
               2
                                302.
                                             33.9
                      173
3
               3
                                             28.4
                      190
                                289.
4
               4
                      185
                                277.
                                             22.7
5
               5
                      198
                                253.
                                             31.8
                                251.
6
               6
                      158
                                             31.2
install.packages("devtools")
devtools::install_github("moodymudskipper/cutr")
install.packages("cutr")
datanew2$quantile <- smart_cut(datanew2$score, 4, "g", output = "numeric")</pre>
datanew2$quantile <- replace(datanew2$quantile, datanew2$quantile==1, "0%-25%")
datanew2$quantile <- replace(datanew2$quantile, datanew2$quantile==2, "25%-50%")
datanew2$quantile <- replace(datanew2$quantile, datanew2$quantile==3, "50%-75%")
datanew2$quantile <- replace(datanew2$quantile, datanew2$quantile==4, "75%-100%")
datanew2 %>%
 group_by(quantile) %>%
 summarise(cutoff=min(score),quality
                                            mean(score),distance=mean(distance))
# A tibble: 4 x 4
   quantile cutoff quality distance
                               <db7>
                                             \langle db 1 \rangle
   <chr>
                   <int>
1 0%-25%
                      158
                                 237.
                                               25.7
                                               28.4
2 25%-50%
                      256
                                 272.
3 50%-75%
                      289
                                 308.
                                               31.2
4 75%-100%
                      330
                                 366.
                                               38.5
# Part2
rm(list = ls())
#Exercise 5-----
set.seed(123)
x1 <- runif(10000, min = 1, max = 3)
x1 <- as.matrix(x1)
x2 < - rgamma(10000, shape = 3, rate = 1/2)
```

```
x2 \leftarrow as.matrix(x2)
x3 <- rbinom(10000, 1, 0.3)
x3 <- as.matrix(x3)
epsilon <- rnorm(10000, mean=2, sd=1)
epsilon <- as.matrix(epsilon)
y < -0.5 + 1.2*x1 - 0.9*x2 + 0.1*x3 + epsilon
ydum <- y
for (i in 1:10000){
  ydum[i] <- 0
  if (y[i]>mean(y)){
    ydum[i] <- 1
  }
}
databook <- data.frame(cbind(y,ydum,x1,x2,x3,epsilon))
names(databook)[names(databook) == "X1"] <- "y"
names(databook)[names(databook) == "X2"] <- "ydum"</pre>
names(databook)[names(databook) == "X3"] <- "x1"
names(databook)[names(databook) == "X4"] <- "x2"
names(databook)[names(databook) == "X5"] <- "x3"
names(databook)[names(databook) == "X6"] <- "epsilon"
#6.1
cor(y,x1)
# Correlation between x1 and y is about 0.20, which very different from 1.2.
> cor(y,x1)
[,1]
 [1,] 0.216015
#6.2$6.3$6.4 Regression of Y on X
cons < -rep(1,10000)
X \leftarrow cbind(cons,x1,x2,x3)
beta <- solve(t(X)%*%X)%*%t(X)%*%y
rownames(beta)[1] <- 'intercept'</pre>
colnames(beta)[1] <- 'est_beta'
sigma2 <- sum((y-X%*%beta)^2)/(nrow(X)-ncol(X))
var <- sigma2*solve(t(X)%*%X)</pre>
SE_ols <- sqrt(diag(var))
SE_ols
```

```
> SE_ols
            cons
0.040620200 0.017358550 0.002876599 0.021694530
#Exercise 7-----
X < -cbind(1,x1,x2,x3)
y <- as.matrix(y)
probit_loglikelihood <- function(b., y. = ydum, X. = X){</pre>
  phi <- pnorm(X.%*%b.)
  phi[phi==1] <- 0.9999 # avoid NaN of log function
  phi[phi==0] <- 0.0001
  f <- sum(y.*log(phi)) + sum((1-y.)*log(1-phi))
  f <- -f
  return(f)
}
probit <- optim(par = c(0,0,0,0), probit_loglikelihood)
probit$par
> probit$par
                           1.17216701 -0.90555423 -0.01106539
        3.04344256
\lceil 1 \rceil
# Optimizing Logit
logit_loglikelihood <- function(b., y. = ydum,X. = X){</pre>
  gamma <- plogis(X%*%b.)
  f <- sum(y.*log(gamma)) + sum((1-y.)*log(1-gamma))
  f <- -f
  return(f)
logit <- optim(par = c(0,0,0,0), logit_loglikelihood)
logit$par
 > logit$par
       5.42762617 2.10006305 -1.61854304 -0.01973273
 Г17
# Optimizing Linear
linear <- lm(ydum \sim x1 + x2 + x3)
summary(linear)
linear$par = c(0.8858236, 0.1461940, -0.1028320, -0.0080531)
linear$par
> linear$par
         0.8858236
                           0.1461940 -0.1028320 -0.0080531
estimation <- cbind(probit$par, logit$par, linear$par)</pre>
colnames(estimation) <- c("Probit", "Logit", "Linear")</pre>
rownames(estimation) <- c("intercept", "x1", "x2", "x3")</pre>
```

^	Probit <sup>‡</sup>	Logit <sup>‡</sup>	Linear <sup>‡</sup>			
intercept	3.04344256	5.42762617	0.8858236			
<b>x1</b>	1.17216701	2.10006305	0.1461940			
x2	-0.90555423	-1.61854304	-0.1028320			
х3	-0.01106539	-0.01973273	-0.0080531			

```
#Exercise 8-----
#Marginal Effect-probit
probit_ME <- function(df){</pre>
  result <- glm(ydum \sim x1 + x2 + x3, family=binomial(link = "probit"),df)
  ME <- mean(dnorm(X%*%coef(result)))*coef(result)
  return(ME)
}
probit_ME(databook)
 > probit_ME(databook)
 (Intercept)
                                     x1
                       0.14380827 -0.11106954 -0.00137997
  0.37324175
#Marginal Effect-logit
logit_ME <- function(df){</pre>
  result <- glm(ydum \sim x1 + x2 + x3, family=binomial(link = "logit"),df)
  ME <- mean(dlogis(X%*%coef(result)))*coef(result)
  return(ME)
}
logit_ME(databook)
> logit_ME(databook)
  (Intercept)
                                     x1
                       0.144030901 -0.110975755 -0.001345977
  0.372080184
#SE of Probit and Logit Marginal Effect by using bootstrapse
bootstrapse <- function(n,fun){</pre>
  boot_result <- data.frame(result = NA)[-1] #creating empty data.frame
  for (i in 1:n) {
    df.existing <- databook[sample(nrow(databook),size = nrow(databook),replace = T),]</pre>
    boot_result <- cbind(boot_result,fun(df.existing))</pre>
  return(data.frame(SE = apply(boot_result,1,sd)))
}
bootstrapse(49,probit_ME)
bootstrapse(49,logit_ME)
```

## > bootstrapse(49,probit\_ME) > bootstrapse(49,logit\_ME)

	SE		SE
(Intercept)	0.0097193201	(Intercept)	0.0100119341
x1	0.0045815902	x1	0.0047115112
x2	0.0005152664	x2	0.0003874686
x3	0.0062492739	x3	0.0068950258