A2

Yue Han

3/24/2021

Exercise 1 Data Description

Average and dispersion in product characteristics

```
apply(margarine$choicePrice[,3:12],2,mean)
```

```
##   PPk_Stk   PBB_Stk   PFl_Stk  PHse_Stk  PGen_Stk  PImp_Stk   PSS_Tub   PPk_Tub
## 0.5184362 0.5432103 1.0150201 0.4371477 0.3452819 0.7807785 0.8250895 1.0774094
##   PFl_Tub  PHse_Tub
## 1.1893758 0.5686734
```

```
apply(margarine$choicePrice[,3:12],2,sd)
```

```
##    PPk_Stk    PBB_Stk    PFl_Stk   PHse_Stk   PGen_Stk   PImp_Stk    PSS_Tub
## 0.15051740 0.12033186 0.04289519 0.11883123 0.03516605 0.11464607 0.06121159
##    PPk_Tub    PFl_Tub   PHse_Tub
## 0.02972613 0.01405451 0.07245500
```

Market share, and market share by product characteristics

```
# Market share
table(margarine$choicePrice$choice)/length(margarine$choicePrice$choice)
```

```
##
##          1          2          3          4          5          6          7
## 0.39507830 0.15637584 0.05436242 0.13266219 0.07046980 0.01655481 0.07136465
##          8          9         10
## 0.04541387 0.05033557 0.00738255
```

```
# Market share by product characteristics
choice_by_price <- t(apply(margarine$choicePrice[,3:12], 1,function(x) x > apply(margarine$choicePrice[

choice_by_price_ <- data.frame(cbind(margarine$choicePrice[,2], choice_by_price))
colnames(choice_by_price_) <- c("choice",1:10)
choice_by_price__ <- choice_by_price_ %>%
  pivot_longer(!choice, names_to = "choice_", values_to = "over_avg") %>%
  filter(choice == choice_) %>%
  select(choice, over_avg)

# Market share whose price under average price
t(table(choice_by_price__))[1,]/length(margarine$choicePrice$choice)
```

```
##           1           2           3           4           5           6
## 0.218791946 0.097539150 0.042505593 0.066442953 0.039149888 0.012527964
```

```
##          7          8          9         10
## 0.026621924 0.019463087 0.005592841 0.003579418
```

```r
# Market share whose price over average price
t(table(choice_by_price__))[2,]/length(margarine$choicePrice$choice)
```

```
##          1          2          3          4          5          6
## 0.176286353 0.058836689 0.011856823 0.066219239 0.031319911 0.004026846
##          7          8          9         10
## 0.044742729 0.025950783 0.044742729 0.003803132
```

Mapping between observed attributes and choices

```r
choice_demos <- merge(margarine$choicePrice, margarine$demos, "hhid")

# mapping between income and choice
table(choice_demos[,c(2,13)])
```

```
##       Income
## choice 2.5 7.5 12.5 17.5 22.5 27.5 32.5 37.5 42.5 47.5  55 67.5 87.5 130
##     1   19 117  196  318  292  195  209  132  125   83  47   19    9   5
##     2    4  54  106  100  123   94   84   34   33   22  30    4   10   1
##     3    0  13   41   27   34    9   28   17   33   23  11    1    3   3
##     4    2  34   44  111  154   67   64   29   23   16  32    8    1   8
##     5    6  19   23   21  123   18   54   23    6    7   7    6    0   2
##     6    0   2    9    5    2    6    4    1   20   17   3    2    1   2
##     7   16  27   40   54   41   24   49   15   27    6  12    7    1   0
##     8    1   6    8   19   36   25   19   14   21    9  42    3    0   0
##     9    2  22   25   20   30   34   33    9   14    2  17    0   12   5
##     10   0   1    3    2    8    4    5    5    1    3   0    1    0   0
```

```r
# mapping between family size and choice
table(choice_demos[,c(2,14)])
```

```
##       Fs3_4
## choice   0   1
##     1  864 902
##     2  339 360
##     3  181  62
##     4  295 298
##     5  128 187
##     6   56  18
##     7  162 157
##     8   81 122
##     9  157  68
##     10  21  12
```

```r
table(choice_demos[,c(2,15)])
```

```
##       Fs5.
## choice    0    1
##     1  1524  242
##     2   621   78
##     3   223   20
##     4   475  118
##     5   252   63
```

```
##     6    51   23
##     7   299   20
##     8   192   11
##     9   214   11
##     10   15   18
```

```r
table(choice_demos[,c(2,16)])
```

```
##       Fam_Size
## choice   1   2   3   4   5   6   7   8
##     1  148 474 400 502 160  76   1   5
##     2   49 212 165 195  53  22   1   2
##     3   38 123  29  33  20   0   0   0
##     4   23 154 119 179  72  33   8   5
##     5   10  55  60 127  33  24   2   4
##     6    7  26  11   7  23   0   0   0
##     7   25 117  77  80   8  12   0   0
##     8   18  52  46  76   2   9   0   0
##     9   34 112  48  20  11   0   0   0
##     10   0   3   3   9  13   5   0   0
```

```r
# mapping between education status and choice
table(choice_demos[,c(2,17)])
```

```
##       college
## choice    0    1
##     1  1205  561
##     2   480  219
##     3   133  110
##     4   419  174
##     5   229   86
##     6    42   32
##     7   216  103
##     8   151   52
##     9   163   62
##     10   18   15
```

```r
# mapping between job status and choice
table(choice_demos[,c(2,18)])
```

```
##       whtcollar
## choice    0    1
##     1   759 1007
##     2   319  380
##     3   111  132
##     4   242  351
##     5    90  225
##     6    32   42
##     7   135  184
##     8    87  116
##     9    95  130
##     10    2   31
```

```r
# mapping between retirement status and choice
table(choice_demos[,c(2,19)])
```

```
##       retired
```

```
## choice    0    1
##      1 1414  352
##      2  531  168
##      3  114  129
##      4  502   91
##      5  269   46
##      6   46   28
##      7  272   47
##      8  183   20
##      9  144   81
##     10   29    4
```

Exercise 2 First Model

Our first model specification is conditional logit model, since the regressors(price) vary across alternatives.

To be specific, we denote p_{ij} as probability of ith individual whose choice is j. $p_{ij} = \frac{\exp(\alpha + x_{ij}\beta)}{\sum_{k=1}^{m}\exp(\alpha + x_{ik}\beta)}$ where $\alpha = [\alpha_1, ..., \alpha_{10}]$. We set $\alpha_1 = 0$.

The negative log likelihood is $-\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}\ln(p_{ij})$

```
choice <- 1:10
names(choice) <- 1:10
y <- as.matrix(map_df(choice, function(x) as.integer(margarine$choicePrice$choice == x)))
x_1 <- choice_demos[,3:12]
```

```
cl_p <- function(x,b) {
    e <- exp(matrix(rep(c(0,b[1:9]),nrow(x)),byrow = TRUE,nrow(x))+x*b[10])
    e_sum <- apply(e,1,sum)
    return(e/e_sum)
}
cl_ll <- function(y,x,b) {
  ln_p <- log(cl_p(x,b))
  return(-sum(y * ln_p))
}
```

```
set.seed(1)
cl <- optim(function(b) cl_ll(y=y,x=x_1,b=b), par = runif(10),method = "BFGS")
cl$par
```

```
##  [1] -0.9543112  1.2970558 -1.7173680 -2.9039702 -1.5152338  0.2518026
##  [7]  1.4649499  2.3575863 -3.8962463 -6.6566826
```

The first 9 parameters are intercepts of goods 2 to 10, and the last parameter is the effect of price.

If one good's intercept is positive, it means that compare to the good 1, individual is more likely to choose that good. One the other hand, if one good's intercept is negative, then individual is less likely to choose that good compare to good 1.

The negative sign of last parameter indicates that the higher the price is, the less likely individual will choose that good.

Exercise 3 Second Model

Our second model specification is multinomial logit model, since the regressors(family income) are invariant across alternatives.

To be specific, we denote p_{ij} as probability of ith individual whose choice is j. $p_{ij} = \frac{\exp(\alpha + x_{ij}\beta_j)}{\sum_{k=1}^{m}\exp(\alpha + x_{ik}\beta_k)}$ where $\alpha = [\alpha_1, ..., \alpha_{10}]$. We set $\alpha_1 = 0$ and $\beta_1 = 0$.

The negative log likelihood is $-\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}\ln(p_{ij})$

```
x_2 <- as.matrix(choice_demos[,13],ncol=1)
```

```
ml_p <- function(x,b) {
  e <- exp(
    matrix(rep(c(0,b[1:9]),nrow(x)),
           byrow = TRUE,
           nrow(x)
           )
    +t(apply(x,1,function(x)x*c(0,b[10:18])))
    )
  e_sum <- apply(e,1,sum)
  return(e/e_sum)
}
ml_ll <- function(y,x,b) {
  ln_p <- log(ml_p(x,b))
  return(-sum(y * ln_p))
}
```

```
set.seed(1)
ml <- optim(function(b) ml_ll(y=y,x=x_2,b=b), par = runif(18), method = "BFGS")
ml$par
```

```
##  [1] -0.843580816 -2.397781498 -1.199508531 -1.688582861 -4.136708740
##  [6] -1.529089710 -2.846092390 -2.573316388 -4.280464931 -0.003154223
## [11]  0.014512648  0.003982003 -0.001326939  0.030526004 -0.007005903
## [16]  0.022809932  0.017664667  0.010709842
```

The first 9 parameters are intercepts of goods 2 to 10, and the last 9 parameters are the effect of income of goods 2 to 10.

If one good's sign of effect is positive, it means that individual with higher income is more likely to choose that good compare to good 1. One the other hand, if one good's sign of effect is negative, individual with higher income is less likely to choose that good compare to good 1.

Exercise 4 Marginal Effects

first model

```
# prob of i individual choose j
p_1 <- cl_p(x_1,cl$par)
# indicator variable
idc <- array(0, dim = c(nrow(x_1),10,10))
for (i in 1:nrow(x_1)) {
  diag(idc[i,,]) <- 1
}
```

```
cl_me <- array(0, dim = c(nrow(x_1),10,10))
for (i in 1:nrow(x_1)) {
  for (j in 1:10) {
    for (k in 1:10) {
      cl_me[i,j,k] <- p_1[i,j]*(idc[i,j,k] - p_1[i,k])*cl$par[10]
    }
```

```
  }
}
```

```
apply(cl_me, c(2,3), mean)
```

```
##               [,1]         [,2]         [,3]         [,4]         [,5]
##  [1,] -1.28527582  0.295366633  0.120714537  0.295075653  0.156234126
##  [2,]  0.29536663 -0.745429822  0.055081150  0.133449019  0.072827819
##  [3,]  0.12071454  0.055081150 -0.337465238  0.050544389  0.030283637
##  [4,]  0.29507565  0.133449019  0.050544389 -0.712655394  0.064017563
##  [5,]  0.15623413  0.072827819  0.030283637  0.064017563 -0.428104692
##  [6,]  0.03732228  0.016726668  0.007105259  0.016551497  0.008749544
##  [7,]  0.15359594  0.069270896  0.029269732  0.063742649  0.037950075
##  [8,]  0.09929634  0.045207285  0.019665695  0.039261859  0.025091839
##  [9,]  0.11082126  0.050699889  0.021755268  0.044153329  0.028521592
## [10,]  0.01684905  0.006800463  0.003045570  0.005859439  0.004428498
##                 [,6]         [,7]         [,8]         [,9]         [,10]
##  [1,]  0.0373222793  0.153595939  0.099296343  0.110821259  0.0168490527
##  [2,]  0.0167266680  0.069270896  0.045207285  0.050699889  0.0068004629
##  [3,]  0.0071052592  0.029269732  0.019665695  0.021755268  0.0030455699
##  [4,]  0.0165514968  0.063742649  0.039261859  0.044153329  0.0058594389
##  [5,]  0.0087495441  0.037950075  0.025091839  0.028521592  0.0044284978
##  [6,] -0.1073284926  0.008538275  0.005430585  0.006113949  0.0007904363
##  [7,]  0.0085382748 -0.420298927  0.025793842  0.027922083  0.0042154367
##  [8,]  0.0054305854  0.025793842 -0.282472167  0.019790097  0.0029346221
##  [9,]  0.0061139488  0.027922083  0.019790097 -0.313060738  0.0032832740
## [10,]  0.0007904363  0.004215437  0.002934622  0.003283274 -0.0482067912
```

It is not surprise that all the diagonal elements are negative while the other elements are all positive. It means that if one good's price increase, people will be willing to choose other goods.

second model

```
# prob of i individual choose j
p_2 <- ml_p(x_2,ml$par)
# beta
ml_b <- c(0,ml$par[10:18])
```

```
ml_me <- array(0, dim = c(nrow(x_2),10))
for (i in 1:nrow(x_2)) {
  b_bar <- sum(p_2[i,]*ml_b)
  for (j in 1:10) {
    ml_me[i,j] <- p_2[i,j]*(ml_b[j]-b_bar)
  }
}
for (i in 1:nrow(x_2)) {
  b_bar <- sum(p_2[i,]*ml_b)
  ml_me[i,] <- p_2[i,]*(ml_b-b_bar)
}
```

```
apply(ml_me, 2, mean)
```

```
##  [1] -1.050861e-03 -9.014940e-04  6.269253e-04  1.660997e-04 -2.794622e-04
##  [6]  4.432084e-04 -6.824357e-04  8.862398e-04  7.339726e-04  5.780756e-05
```

For goods 1, 2, 5, 7, if individual's income raise, he or she will choice these goods less, and will turn to the rest goods.

Exercise 5 IIA

```
mx_ll <- function(y,x,b,mx_p) {
  ln_p <- log(mx_p(x,b))
  return(-sum(y * ln_p))
}
```

full data

We denote p_{ij} as probability of ith individual whose choice is j. $p_{ij} = \frac{\exp(\alpha+x_{ij}\beta+w_i\gamma_j)}{\sum_{k=1}^{m}\exp(\alpha+x_{ik}\beta++w_i\gamma_k)}$ where $\alpha = [\alpha_1, ..., \alpha_{10}]$. We set $\alpha_1 = 0$ and $\gamma_1 = 0$.

The negative log likelihood is $-\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij}\ln(p_{ij})$

```
x_3 <- as.matrix(choice_demos[,3:13],ncol=1)
```

```
mx_p_1 <- function(x,b) {
    e <- exp(
    matrix(rep(c(0,b[1:9]),nrow(x)),
           byrow = TRUE,
           nrow(x)
           )
    +x[,1:10]*b[10]
    +t(apply(matrix(x[,11],ncol=1),1,function(x)x*c(0,b[11:19])))
    )
    e_sum <- apply(e,1,sum)
    return(e/e_sum)
}
```

```
set.seed(1)
mx_1 <- optim(function(b) mx_ll(y=y,x=x_3,b=b,mx_p=mx_p_1), par = runif(19), method = "BFGS")
```

β^f:

```
mx_1$par
```

```
##  [1] -0.838736865  0.890962176 -1.826473674 -2.871207153 -2.454392513
##  [6]  0.498906843  0.805552527  1.866518262 -4.139801155 -6.659575193
## [11] -0.004332036  0.014258494  0.004025558 -0.001264376  0.029719239
## [16] -0.009323928  0.021909429  0.016904583  0.008667871
```

remove second choice

```
x_4 <- x_3[,-2]
```

```
mx_p_2 <- function(x,b) {
    e <- exp(
    matrix(rep(c(0,b[1:8]),nrow(x)),
           byrow = TRUE,
           nrow(x)
           )
    +x[,1:9]*b[9]
    +t(apply(matrix(x[,10],ncol=1),1,function(x)x*c(0,b[10:17])))
    )
```

```r
    e_sum <- apply(e,1,sum)
    return(e/e_sum)
}
```

```r
set.seed(1)
mx_2 <- optim(function(b) mx_ll(y=y[,-2],x=x_4,b=b,mx_p=mx_p_2), par = runif(17), method = "BFGS")
```

β^r:

```r
mx_2$par
```

```
##  [1]  0.546387162 -1.710865690 -2.685532157 -2.608530870  0.283032337
##  [6]  0.392558598  1.375902478 -4.132033516 -5.881445504  0.014219308
## [11]  0.003845320 -0.001604402  0.029823641 -0.009242461  0.022381420
## [16]  0.017128602  0.009298403
```

MTT statistics

```r
l_1 <- mx_ll(y=y[,-2],x=x_4,b=mx_1$par[-c(1,11)],mx_p=mx_p_2)
l_2 <- mx_ll(y=y[,-2],x=x_4,b=mx_2$par,mx_p=mx_p_2)
MTT <- 2*(l_1-l_2)
c_v <- qchisq(0.95, length(mx_2$par))
MTT < c_v
```

```
## [1] TRUE
```

Since MTT less than critical value, we conclude that under 5% significant level, we cannot reject that IIA hold.