Assignment 3

Xiaoshu Gui

Setup

```r
library(dplyr)
library(purrr)
library(magrittr)
library(tidyr)
library(tibble)
library(stringr)
library(reshape2)
library(lme4)
library(mfx)
library(bayesm)
library(mclogit)
library(mlogit)
library(nnet)
# load data
data(margarine)
# Create a dataframe that merges product characteristics with household demos by hhid.
choiceprice <- as.matrix(margarine$choicePrice)
demos <- as.matrix(margarine$demos)
marg <- merge(choiceprice, demos, by = "hhid")
```

Exercise 1 Data Description

- Average and dispersion in product characteristics.

```r
# average:
apply(marg[, 3:12], 2, mean)
```

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

```r
# dispersion
apply(marg[, 3:12], 2, sd)
```

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

- Market share, and market share by product characteristics.

```r
# market share by product
ms_product <- table(marg$choice)/4470
names(ms_product) <- names(marg[,3:12])
print(ms_product)
```

```
##    PPk_Stk    PBB_Stk    PFl_Stk   PHse_Stk   PGen_Stk   PImp_Stk
## 0.39507830 0.15637584 0.05436242 0.13266219 0.07046980 0.01655481
##    PSS_Tub    PPk_Tub    PFl_Tub   PHse_Tub
## 0.07136465 0.04541387 0.05033557 0.00738255
```

```r
# market share by product characteristics: brand and type
brand_name <- names(marg[,3:12]) %>%
  str_replace_all("_Stk|_Tub", "")
ms_brand <- cbind.data.frame(brand_name, ms_product) %>%
  group_by(brand_name) %>%
  summarise(market_share = sum(Freq))
print(ms_brand)
```

```
## # A tibble: 7 x 2
##   brand_name market_share
##   <fct>             <dbl>
## 1 PBB              0.156
## 2 PFl              0.105
## 3 PGen             0.0705
## 4 PHse             0.140
## 5 PImp             0.0166
## 6 PPk              0.440
## 7 PSS              0.0714
```

```r
# by product type (stick and tub)
sum(ms_product[1:6]) # market share of stick
```

```
## [1] 0.8255034
```

```r
sum(ms_product[7:10]) # market share of tub
```

```
## [1] 0.1744966
```

- Mapping between observed attributes and choices.

Create tables of choices by different household attributes:

```r
# income level & choices
table(marg$Income, marg$choice)
```

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

```r
# family size:
table(marg$Fam_Size, marg$choice)
```

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

```
##    7    1    1    0    8    2    0    0    0    0    0
##    8    5    2    0    5    4    0    0    0    0    0
```

```r
# education status & choices
table(marg$college, marg$choice)
```

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

```r
# job status & choicies
table(marg$whtcollar, marg$choice)
```

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

```r
# retirement status & choices
table(marg$retired, marg$choice)
```

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

Recap Multinomial Models

There are m alternatives and the dependent variable y is defined to take value j if the jth alternative is taken, j = 1, . . . , m. Based on the random utility model, let U_{ij} denote the utility of individual i derive when choosing altertive j. j is chosen if and only if $U_{ij} > U_{ik}$ for all $k \neq j$. Although we can't observe U_{ij}, we can treat it as independent random variables with a systematic component V_{ij} and a random component ϵ_{ij} such that $U_{ij} = V_{ij} + \epsilon_{ij}$.

Define the probability that alternative j is chosen by individual i as:

$$P_{ij} = Pr[y_i = j] = \frac{V_{ij}}{\sum_{k=1}^{m} V_{ik}}, \quad j = 1, ..., m,$$

where $V_{ij} > 0$ can be general functions of regressors X_i and parameters β. This is a *universal logit model*. Different specifications for V_{ij} corresponds to specific models, such as multinomial logit and conditional logit models. In that sense, all these models are variants of the same model. They only differ in their parametrization of the systematic compotents V_{ij}.

The log likelihood function of the universal logit model is:

$$L = \sum_{i=1}^{N} \sum_{j=1}^{m} y_{ij} \; ln \; P_{ij}, \qquad j = 1, ..., m, \quad i = 1, ..., N,$$

where P_{ij} is defined above.

Exercise 2 First Model

- We are interested in the effect of price on demand. Propose a model specification.

Since the price of a product varies by different choices, a conditional logit model is chosen to deal with regressors varying across alternatives. Here $V_{ij} = X_{ij}$, specifying characteristics of the alternatives (price).

The probability of the ith househould choosing product j is given by

$$P_{ij} = Pr[y_i = j] = \frac{\exp(X_{ij}\beta)}{\sum_{k=1}^{m} exp(X_{ik}\beta)}, \quad j = 1, ...m$$

where X denotes price, the substrcipt i denotes the ith household, subscript j or k denotes the alternative, and parameter β is contant across alternatives. Note that it is possible to go from alternative-varying regressors to alternative-invariant format. Let X_i be a $K * 1$ vector. Define X_{ij} to be a $Km * 1$ vector with zeros except that the jth block is X_i, that is $X_{ij} = [0'...0', X_i, 0', ...0']'$, and define

$$\beta = [0', \beta_2', ...\beta_m']'$$

, where $\beta_1 = 0$ is a normalization. Then $X_i'\beta_j = X_{ij}'\beta$.

The likelihood function of conditional multinomial model is:

```
loglik_cl = function(beta) {
  X = marg[, 3:12] - marg[, 3] # set price of the first product as reference
  b = beta[1] # constant beta
  alpha = beta[2:11]
  alpha[1] = 0
  X_beta = X * b
  alpha_choice = matrix(nrow = nrow(marg), ncol = 1)
  X_beta_j = matrix(nrow = nrow(marg), ncol = 1)
  alpha_t = matrix(rep(t(alpha), times = nrow(marg)), ncol = ncol(t(alpha)), byrow = T)
  for (i in 1: nrow(marg)) {
    jstar  = marg[i, "choice"]
    alpha_j = alpha[jstar]
    alpha_choice[i] = alpha_j
    X_beta_j[i] = X_beta[i, jstar]
  }
  numerator = exp(X_beta_j + alpha_choice)
  Xbeta_k =  exp(X_beta + alpha_t)
  denominator = rowSums(Xbeta_k)
  Pij = numerator / denominator
  ll = log(Pij)
  loglik_cl = - sum(ll)
}
# optimize the likelihood using the nlm function.
fit_cl = nlm(f = loglik_cl, p = c(rep(0, 11)))
fit_cl$estimate
```

```
##  [1] -6.6565906  0.0000000 -0.9543061  1.2969786 -1.7173341 -2.9040074
##  [7] -1.5153149  0.2517637  1.4648579  2.3575174 -3.8965934
```

```
#
#a <- function(beta) {
#  X = marg[, 3:12] ## price takes dif values for dif alternatives
#  beta = matrix(nrow = nrow(marg), ncol = 1)
#  X_beta = X * beta # a matrix of 4479*10
#  Xj_beta = matrix(nrow = nrow(marg), ncol = 1) # 4470*1
#  for (i in 1: nrow(marg)) {
#    jstar = marg[i, "choice"] # choice made by household i.
    # beta_i = beta[i] # the effect of price is the same across alternatives but different for decision maker
#    Xj_beta[i] = X_beta[i, jstar] # define X_ij*beta
#  }
#  numerator = exp(Xj_beta)
#  denominator = rowSums(X_beta)
#  Pij = numerator/denominator
#  ll = log(Pij)
#  a = - sum(ll)
#}

#fit_a = nlm(f = a, p = c(rep(0, 1)))
#fit_a
```

Interpret the coefficient on price: Note that the constant beta < 0, suggesting that an increase in the price of one alternative decreases the probability of choosing that alternative and increases the probability of choosing other alternatives.

Exercise 3 Second Model

We are interested in the effect of family income on demand. Propose a model specification.

Since family income is a fixed constant for decision makers (households) and does not vary across product choices, a multinomial logit model is chosen to address alternative-invariant regressors. The probability of the ith household choosing product j is given by:

$$P_{ij} = Pr[y_i = j] = \frac{\exp(\alpha_j + X_i\beta_j)}{\sum_{k=1}^{m} exp(\alpha_k + X_i\beta_k)}, \quad j = 1, ...m,$$

where X denotes income. The likelihood function is:

```
loglik_mnl = function(beta) {
  X = as.matrix(cbind(marg[, 13:15], marg[, 17:19], rep(1, nrow(marg)))) # 4470*7
  beta = matrix(nrow = 7, byrow = T) # create a matrix of 7*1
  X_beta_j = X %*% beta # 4470*1

  for (i in 1:nrow(marg)) {
  X_beta_j[i] = X[i, ] %>% beta
  }
  numerator = exp(X_beta_j)
  denominator = sum(exp(X_beta_j))
  Pij = numerator/denominatorn
  ll = log(Pij)
  loglik_mnl = - sum(ll)
}
# optimize the likelihood of multinomial model:
#fit_mnl = nlm(f = loglik_mnl, p = rep(0, 7)) # something is wrong here, fail to debug.
#fit_mnl
```

Exercise 4 Marginal Effects

Compute and interpret the marginal effects for the first and second models.

```
## Marginal effect of the conditional logit model
X = marg[, 3:12] - marg[, 3]
b = fit_cl$estimate[1]
alpha = fit_cl$estimate[2:11]
X_beta = X * b
alpha_choice = matrix(nrow = nrow(marg), ncol = 1)
X_beta_j = matrix(nrow = nrow(marg), ncol = 1)
alpha_t = matrix(rep(t(alpha), times = nrow(marg)), ncol = ncol(t(alpha)), byrow = T)
xbetak =  exp(X_beta + alpha_t)
denominator = rowSums(xbetak)
pr_ij = as.matrix(xbetak/denominator)
pij = t(pr_ij) %*% pr_ij * (-b)
a = matrix(rep(colSums(pr_ij) * b,10), ncol=10 )
a = a * diag(10)
me_cl = data.frame((pij + a)/nrow(marg))
me_cl
```

```
##                PPk_Stk      PBB_Stk      PFl_Stk    PHse_Stk     PGen_Stk
## PPk_Stk    -1.28526906  0.295370795  0.120711900  0.29508412  0.156227495
## PBB_Stk     0.29537079 -0.745429040  0.055079933  0.13345281  0.072824647
## PFl_Stk     0.12071190  0.055079933 -0.337453813  0.05054413  0.030281218
## PHse_Stk    0.29508412  0.133452809  0.050544129 -0.71266485  0.064015927
## PGen_Stk    0.15622750  0.072824647  0.030281218  0.06401593 -0.428082220
## PImp_Stk    0.03732038  0.016725820  0.007104638  0.01655091  0.008748605
## PSS_Tub     0.15359586  0.069270843  0.029268647  0.06374367  0.037947887
## PPk_Tub     0.09929335  0.045205906  0.019664358  0.03926138  0.025089627
```

```
## PFl_Tub   0.11082171  0.050700063  0.021754537  0.04415429  0.028520040
## PHse_Tub  0.01684346  0.006798224  0.003044452  0.00585762  0.004426773
##               PImp_Stk      PSS_Tub      PPk_Tub      PFl_Tub
## PPk_Stk   0.0373203777  0.153595860  0.099293347  0.110821713
## PBB_Stk   0.0167258197  0.069270843  0.045205906  0.050700063
## PFl_Stk   0.0071046380  0.029268647  0.019664358  0.021754537
## PHse_Stk  0.0165509100  0.063743674  0.039261382  0.044154286
## PGen_Stk  0.0087486052  0.037947887  0.025089627  0.028520040
## PImp_Stk -0.1073218508  0.008537721  0.005430073  0.006113580
## PSS_Tub   0.0085377214 -0.420292978  0.025792624  0.027921746
## PPk_Tub   0.0054300734  0.025792624 -0.282460043  0.019789215
## PFl_Tub   0.0061135797  0.027921746  0.019789215 -0.313057324
## PHse_Tub  0.0007901258  0.004213974  0.002933510  0.003282144
##               PHse_Tub
## PPk_Stk   0.0168434590
## PBB_Stk   0.0067982244
## PFl_Stk   0.0030444522
## PHse_Stk  0.0058576197
## PGen_Stk  0.0044267729
## PImp_Stk  0.0007901258
## PSS_Tub   0.0042139745
## PPk_Tub   0.0029335105
## PFl_Tub   0.0032821436
## PHse_Tub -0.0481902824
```

```r
## Marginal effect of the multinomial model
X = as.matrix(cbind(marg[, 13:15], marg[, 17:19], rep(1, nrow(marg))))
#beta = matrix(fit_mnl$estimate, nrow = 7, byrow = T)
#X_beta_j = X %*% beta
#ex = exp(X_beta_j)
#Pij = t((apply(ex, 1, function(x) x / sum(x))))
#beta_income = matrix(beta[1, ])
#beta_bar = Pij %*% beta_income
#beta_bar_large = matrix(rep(beta_bar, 10), ncol = 10)
#beta_j = matrix(rep(t(beta_income)), nrow(data), byrow = T, ncol = 10)
#me_mnl = data.frame(colSums(Pij * (beta_j - beta_bar_large))/nrow(data))
#me_mnl
```

Exercise 5 IIA

Now combine the above two models to estimate the effect of price and family income on choices of margarine. The mixed logit model is specified as:

$$P_{ij} = \frac{exp(X_{ij}\beta \ + W_i\gamma_j)}{\sum_{k=1}^{m} exp(X_{ik}\beta + W_i\gamma_k)}, \quad j = 1, ..., m$$