Choice-based conjoint analysis

Avet Mnatsakanyan

2022-12-24

This project presents choice-based conjoint analysis conducted for yogurt market. Respondents were asked to chose one of the 3 alternative options from 15 different combinations of yogurt product attributes. Atributes are:

- diet - yes/no
- size - 5/6/8 ounces
- shape - circle/square
- flavor - vanilla/cherry/peach
- price - 4/6/9 dollar

Additional variables in dataset are:

- resp_id - unique numerical id for each respondent
- ques - question number asked to each respondent (1:15)
- alt - options presented in each question (1:3)
- choice - 1 for chosen, 0 for not chosen

The cross tabulation below shows choices for each atribute:

```
rm(list=ls())
require(mlogit)
```

```
## Loading required package: mlogit

## Warning: package 'mlogit' was built under R version 4.1.3

## Loading required package: dfidx

## Warning: package 'dfidx' was built under R version 4.1.3

##
## Attaching package: 'dfidx'

## The following object is masked from 'package:stats':
##
##     filter
```

```
setwd("C:\\Users\\Avet\\Desktop\\Git\\R\\Conjoint analysis") ## directory


yogurt = read.csv("conjoint_yogurt.csv")

attach(yogurt)

yogurt$size = as.factor(yogurt$size)
yogurt$price = as.factor(yogurt$price)
```

```r
# cross tabulations
xtabs(choice~ diet, data= yogurt)
```

```
## diet
##   no  yes
## 1830 1170
```

```r
xtabs (choice ~ shp, data=yogurt)
```

```
## shp
## crcl  sqr
## 1709 1291
```

```r
xtabs (choice ~ price, data=yogurt)
```

```
## price
##    4    6    9
## 1466  984  550
```

```r
xtabs (choice ~ size, data=yogurt)
```

```
## size
##    5    6    8
## 1112  881 1007
```

```r
xtabs (choice ~ flav, data=yogurt)
```

```
## flav
## cher peac  van
##  580  982 1438
```

Reshape the dataset using mlogit modelling (multinomial logit model)

```r
# add a column with unique question numbers, as needed in mlogit 1.1+
yogurt$chid <- rep(1:(nrow(yogurt)/3), each=3)
# shape the data for mlogit
yogurt.mlogit<- dfidx(yogurt, choice="choice", idx=list(c("chid", "resp.id"), "alt" ))
```

Then run the logit model without intercept, assuming that the position of the question in the survey doesn't matter for respondents

```r
attach(yogurt.mlogit)
```

```
## The following objects are masked from yogurt:
##
##     choice, diet, flav, price, ques, shp, size
```

```r
m1 <- mlogit(choice ~ 0  + size + shp + flav + price, data = yogurt.mlogit)
summary(m1)
```

```
##
## Call:
## mlogit(formula = choice ~ 0 + size + shp + flav + price, data = yogurt.mlogit,
##     method = "nr")
##
## Frequencies of alternatives:choice
##     1     2     3
```

```
## 0.325 0.347 0.328
##
## nr method
## 5 iterations, 0h:0m:0s
## g'(-H)^-1g = 5.1E-05
## successive function values within tolerance limits
##
## Coefficients :
##           Estimate Std. Error  z-value  Pr(>|z|)
## size6    -0.395743   0.062733  -6.3084  2.82e-10 ***
## size8    -0.153658   0.061339  -2.5051   0.01224 *
## shpsqr   -0.520698   0.050971 -10.2156 < 2.2e-16 ***
## flavpeac  0.784948   0.066954  11.7237 < 2.2e-16 ***
## flavvan   1.601449   0.068335  23.4354 < 2.2e-16 ***
## price6   -0.796060   0.059407 -13.4001 < 2.2e-16 ***
## price9   -1.621879   0.067511 -24.0239 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -2597.6
```

```r
round(exp(coef(m1)),3)
```

```
##    size6    size8   shpsqr flavpeac  flavvan   price6   price9
##    0.673    0.858    0.594    2.192    4.960    0.451    0.198
```

Include intercept and conduct 'lrtest' to bring an argument in favor of assumption provided above

```r
m2 <- mlogit(choice ~  size +shp + flav +price , data = yogurt.mlogit)
summary(m2)
```

```
##
## Call:
## mlogit(formula = choice ~ size + shp + flav + price, data = yogurt.mlogit,
##     method = "nr")
##
## Frequencies of alternatives:choice
##     1     2     3
## 0.325 0.347 0.328
##
## nr method
## 5 iterations, 0h:0m:0s
## g'(-H)^-1g = 5.47E-05
## successive function values within tolerance limits
##
## Coefficients :
##                 Estimate Std. Error  z-value  Pr(>|z|)
## (Intercept):2   0.083989   0.050859   1.6514   0.09865 .
## (Intercept):3   0.032759   0.051417   0.6371   0.52404
## size6          -0.395597   0.062757  -6.3037 2.907e-10 ***
## size8          -0.153251   0.061355  -2.4978   0.01250 *
## shpsqr         -0.521324   0.050982 -10.2256 < 2.2e-16 ***
## flavpeac        0.785964   0.066974  11.7354 < 2.2e-16 ***
## flavvan         1.601473   0.068372  23.4228 < 2.2e-16 ***
```

```
## price6        -0.799471   0.059488 -13.4392 < 2.2e-16 ***
## price9        -1.624457   0.067570 -24.0412 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -2596.2
## McFadden R^2:  0.21196
## Likelihood ratio test : chisq = 1396.6 (p.value = < 2.22e-16)
```

```
lrtest(m1, m2)
```

```
## Likelihood ratio test
##
## Model 1: choice ~ 0 + size + shp + flav + price
## Model 2: choice ~ size + shp + flav + price
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   7 -2597.6
## 2   9 -2596.2  2 2.7794     0.2492
```

Based on Chi-square (0.2492>0.05) two models are statistically the same.

Let's include price variable as numeric instead of treating it as factor:

```
m3 <- mlogit(choice ~   0+ size +shp + flav +as.numeric(price) , data = yogurt.mlogit)
summary(m3)
```

```
##
## Call:
## mlogit(formula = choice ~ 0 + size + shp + flav + as.numeric(price),
##     data = yogurt.mlogit, method = "nr")
##
## Frequencies of alternatives:choice
##     1     2     3
## 0.325 0.347 0.328
##
## nr method
## 5 iterations, 0h:0m:0s
## g'(-H)^-1g = 4.77E-05
## successive function values within tolerance limits
##
## Coefficients :
##                    Estimate Std. Error  z-value  Pr(>|z|)
## size6             -0.395718   0.062739  -6.3074 2.838e-10 ***
## size8             -0.153312   0.061324  -2.5000   0.01242 *
## shpsqr            -0.520833   0.050971 -10.2182 < 2.2e-16 ***
## flavpeac           0.785206   0.066945  11.7291 < 2.2e-16 ***
## flavvan            1.601894   0.068324  23.4457 < 2.2e-16 ***
## as.numeric(price) -0.809854   0.033506 -24.1701 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -2597.7
```

```
lrtest(m1, m3)
```

```
## Likelihood ratio test
```

```
##
## Model 1: choice ~ 0 + size + shp + flav + price
## Model 2: choice ~ 0 + size + shp + flav + as.numeric(price)
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   7 -2597.6
## 2   6 -2597.7 -1 0.0789     0.7787
```

Again, no statistical difference between model 1 and model 3. So, from the output above we can get the answer of some marketing questions:

- On average, consumers prefer Yogurts with size 5 oz over 6 and 8 oz (if we hold other attributes the same)
- Square shapes are less attractive compared to circle shapes
- Flavor of peach and vanilla are more preferable than flavor of cherry
- And , obviously, people like lower prices

However, instead of observing only directions of indicators we can compute the average willingness-to-pay for a particular level of an attribute by dividing the coefficient for that level by the price coefficient:

```
round(coef(m3)[-6]/(-coef(m3)["as.numeric(price)"]),1)
```

```
##    size6    size8   shpsqr flavpeac  flavvan
##     -0.5     -0.2     -0.6      1.0      2.0
```

The final findings from this analysis are following:

- On average, customers would be equally divided between a yogurt with 5oz of size and yogurt with 6 oz of size that costs $0.5 less. Another way to think of it is that $0.5 is the price at which consumers become indifferent between the two sizes. At the same time consumers become indifferent between the 5 oz and 8 oz options when the 8 oz product costs $0.2 less.

- On average , customers would be equally divided based on their preferences for yogurt shape options (circle vs square) when the product with square shape costs $0.6 less.

- Consumers would be indifferent between flavor of cherry and flavor of peach if the latter costs $1 more. At the same time, the indifference would be achieved between flavor of cherry and flavor of vanilla when yogurt with vanilla flavor costs $2 more.