

# Choice Based Conjoint MNL model: Gym Subscriptions Case Study

Code ▾

Among the public places hardest hit by the lockdown to try to bring down the rate of coronavirus, there are certainly gyms, which have been closed since October and will probably be the last business to reopen. In order to understand which are the factors on which one can invest to relaunch the sector, survey participants are asked to choose one of different subscription profiles. The various types of subscription include:

- subscription duration (monthly / three-months / yearly)
- type of activities (only fitness room / fitness room plus courses)
- time slots in which the costumer can access the structure (7-12 / 18-23 / all day)
- type of training schedule (standard / personalized) - additional services (not included / access to the wellness center / physiotherapy consultancy)
- corresponding price per month (40 euros / 60 euros / 80 euros)

The aim of the data analysis is to assess the relationship between the choice and the product attributes. To do this we use a discrete choice model, such as the multinomial logit regression.

Hide

```
palestra <- read.csv("C:/Users/tamar/OneDrive/Desktop/DS-2 anno/lab of costumer/palestra.csv", sep=";")
dim(palestra) #24000 x 10
str(palestra)
```

Hide

```
head(palestra)
```

resp.id	q...	alt	abbonamento	attività	orario	tabella	servizi	
<int>	<int>	<int>	<chr>	<chr>	<chr>	<chr>	<chr>	
1	1	1	1 mensile	attrezzi+corsi	7_12	personalizzata	centro.benessere	
2	1	1	2 mensile	attrezzi+corsi	18_23	standard	consulenza.fisioterapica	
3	1	1	3 annuale	attrezzi+corsi	18_23	standard	centro.benessere	
4	1	1	4 mensile	attrezzi	18_23	personalizzata	centro.benessere	
5	1	2	1 annuale	attrezzi	allDay	personalizzata	centro.benessere	
6	1	2	2 mensile	attrezzi	allDay	personalizzata	non.inclusi	

6 rows | 1-9 of 10 columns

Data are in long format and each respondent has to choose between 4 alternative.

Hide

```
options(max.print=100)
sapply(palestra,table) #distribution of each variable
```

```
$resp.id

 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
80 80 80 80 80 80 80 80 80 80 80 80 80 80 80 80
17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
80 80 80 80 80 80 80 80 80 80 80 80 80 80 80 80
33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48
80 80 80 80 80 80 80 80 80 80 80 80 80 80 80 80
49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64
80 80 80 80 80 80 80 80 80 80 80 80 80 80 80 80
65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
80 80 80 80 80 80 80 80 80 80 80 80 80 80 80 80
81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96
80 80 80 80 80 80 80 80 80 80 80 80 80 80 80 80
97 98 99 100
80 80 80 80

[ reached getOption("max.print") -- omitted 200 entries ]

$ques

 1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4
17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32
4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4
33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48
4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4
49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64
4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4
65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4
81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96
4  4  4  4  4  4  4  4  4  4  4  4  4  4  4  4
97 98 99 100
4  4  4  4

[ reached getOption("max.print") -- omitted 5900 entries ]

$alt

 1  2  3  4
6000 6000 6000 6000

$abbonamento
```

	annuale	mensile	trimestrale
	8023	8014	7963
\$attività			
	attrezzi	attrezzi+corsi	
	12110	11890	
\$orario			
	18_23	7_12	allDay
	8018	8035	7947
\$tabella			
	personalizzata	standard	
	11927	12073	
\$servizi			
	centro.benessere	consulenza.fisioterapica	
		8050	8034
	non.inclusi		
	7916		
\$prezzo			
	40	60	80
	7989	7996	8015
\$choice			
	0	1	
	18000	6000	

We can see that each respondent has look at 80 product profiles. Each level of each attribute is properly represented and well balanced, as all levels have the same frequency. There are no respondent level variables.

See some descriptive statistics:

Hide

summary(palestra)

```

  resp.id      ques      alt
Min.   : 1.00  Min.   : 1  Min.   :1.00
1st Qu.: 75.75 1st Qu.:1501 1st Qu.:1.75
Median :150.50 Median :3000 Median :2.50
Mean   :150.50 Mean   :3000 Mean   :2.50
3rd Qu.:225.25 3rd Qu.:4500 3rd Qu.:3.25
Max.   :300.00 Max.   :6000 Max.   :4.00
abbonamento  attivita  orario
Length:24000   Length:24000 Length:24000
Class :character Class :character Class :character
Mode  :character Mode  :character Mode  :character

  tabella      servizi      prezzo
Length:24000   Length:24000   Min.   :40.00
Class :character Class :character 1st Qu.:40.00
Mode  :character Mode  :character   Median :60.00
                                   Mean    :60.02
                                   3rd Qu.:80.00
                                   Max.    :80.00

  choice
Min.   :0.00
1st Qu.:0.00
Median :0.00
Mean   :0.25
3rd Qu.:0.25
Max.   :1.00

```

Hide

xtabs(choice ~ abbonamento, data=palestra) #On average consumers prefer annual subscription

abbonamento		
annuale	mensile	trimestrale
3006	1688	1306

Hide

xtabs(choice ~ attivita, data=palestra) #On average consumers prefer fitness room plus courses instead of the only fitness room

attivita	
attrezzi	attrezzi+corsi
2536	3464

Hide

xtabs(choice ~ orario, data=palestra)#On average consumers prefer the evening schedule

```
orario
18_23 7_12 allDay
3242 1962 796
```

Hide

```
xtabs(choice ~ tabella, data=palestra) #On average consumers prefer the standard training schedule
```

```
tabella
personalizzata      standard
      2719             3281
```

Hide

```
xtabs(choice ~ servizi, data=palestra) #On average consumers prefer physiotherapy consultancy as an additional service.
```

```
servizi
      centro.benessere consulenza.fisioterapica
              1031              2771
      non.inclusi
              2198
```

Hide

```
xtabs(choice ~ prezzo, data=palestra) #On average consumers prefer the lower price (40 euros/month)
```

```
prezzo
      40  60  80
2312 1897 1791
```

## MULTINOMIAL LOGIT MODEL

Since customers have to choose between 4 alternative, we are in the context of a multilevel logit model.

First of all we have to recode the variable as qualitative, also by changing the categories order of some of them in order to have a specific reference level.

Hide

```
palestra$abbonamento <- factor(palestra$abbonamento, levels = c("mensile","trimestrale","annuale"))
palestra$attivita <- as.factor(palestra$attivita)
palestra$orario<-factor(palestra$orario, levels = c("allDay", "18_23","7_12"))
palestra$tabella<- factor(palestra$tabella, levels=c("standard","personalizzata" ))
palestra$servizi<-factor(palestra$servizi, levels=c("consulenza.fisioterapica",
"centro.benessere", "non.inclusi" ))
palestra$prezzo<-as.factor(palestra$prezzo)
```

Then organize the data in long format in the properly way for the mlogit function:

Hide

```
palestra_mlogit <- dfixd(palestra, idx = list(c("ques", "resp.id"), "alt"), drop.index=F,
      levels=c("1", "2", "3", "4"))
```

Now we can fit the MNL model:

Hide

```
m1 <- mlogit(choice ~ abbonamento +attivita + orario+ tabella + servizi + prezzo, data = palestra_mlogit)
summary(m1)
```

```
Call:
mlogit(formula = choice ~ abbonamento + attivita + orario + tabella +
servizi + prezzo, data = palestra_mlogit, method = "nr")
```

```
Frequencies of alternatives:choice
      1      2      3      4
0.25633 0.25500 0.24150 0.24717
```

```
nr method
5 iterations, 0h:0m:1s
g'(-H)^-1g = 1.82E-07
gradient close to zero
```

```
Coefficients :
              Estimate Std. Error z-value Pr(>|z|)
(Intercept):2      -0.024577   0.043446  -0.5657   0.5716
(Intercept):3      -0.066762   0.043885  -1.5213   0.1282
(Intercept):4      -0.025792   0.043810  -0.5887   0.5560
abbonamentotrimetrale -0.407393   0.047441  -8.5874 <2e-16
abbonamentoannuale  1.146834   0.043888  26.1312 <2e-16
attivitaattrezzi+corsi  0.623608   0.036542  17.0654 <2e-16
orario18_23         2.359177   0.054307  43.4414 <2e-16
orario7_12          1.397186   0.052572  26.5765 <2e-16
tabellapersonalizzata -0.333451   0.035824  -9.3080 <2e-16
servizicentro.benessere -1.756346   0.050374 -34.8662 <2e-16
servizinon.inclusi   -0.438293   0.041126 -10.6572 <2e-16
prezzo60             -0.379502   0.043324  -8.7597 <2e-16
prezzo80             -0.507472   0.044098 -11.5080 <2e-16
```

```
(Intercept):2
(Intercept):3
(Intercept):4
abbonamentotrimestrale ***
abbonamentoannuale ***
attivitaaattrezzi+corsi ***
orario18_23 ***
orario7_12 ***
tabellapersonalizzata ***
servizicentro.benessere ***
servizinon.inclusi ***
prezzo60 ***
prezzo80 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -5676
McFadden R^2: 0.31747
Likelihood ratio test : chisq = 5280.1 (p.value = < 2.22e-16)
```

Considering that estimated coefficients are on the logit scale and tend to range mainly between -2 and 2, we can conclude that on average:

- customers strongly prefer the yearly subscription compared to the monthly one, on the contrary the three-months subscription is slightly less preferred than the monthly one
- customers prefer that in the subscription is included the possibility of attending both fitness room and the courses instead of the only fitness room
- customers strongly prefer a cheaper subscription that allows them to enter the gym at a specific time slot than the more expensive subscription that allows them to enter at any time. The preferred time slot is 18-23
- customers prefer the standard training schedule
- customers are strongly not interested having the in wellness center as an additional service, while the physiotherapy consultancy is preferred over having the additional services not included.
- customers prefers the lower price

Concerning the alternative specific constants, they are not significant and indeed here we are not expected a respondent chooses a subscription option because it is positioned in a certain way. We can check if this intuition is correct by fitting a nested smaller model without intercept parameters and then by comparing it with the larger model with intercepts through a likelihood ratio test.

Hide

```
m2 <- mlogit(choice ~ abbonamento +attivit  + orario+ tabella + servizi + prezzo| -1, data = palestra_mlogit)
lrtest(m2,m1) #comparison between models
```

```
Likelihood ratio test

Model 1: choice ~ abbonamento + attivita + orario + tabella + servizi +
prezzo | -1
Model 2: choice ~ abbonamento + attivita + orario + tabella + servizi +
prezzo
#Df LogLik Df Chisq Pr(>Chisq)
1 10 -5677.1
2 13 -5676.0 3 2.3767 0.498
```

The comparison leads to a large p-value of 0.498. Therefore we can conclude that the two models are not significantly different in terms of goodness of fit and they explain the data equally well. As consequences, alternative specific constants are not necessary to adequately model the data.

## Compute the willingness-to-pay

The likelihood ratio test can be also useful to decide if the price variable must be included as quantitative predictor instead of being qualitative. Therefore we now fit the model without intercept parameters and with *prezzo* as a quantitative variable in order to compare it with the restricted model with *prezzo* as qualitative.

Hide

```
m3 <- mlogit(choice ~ abbonamento +attivit  + orario+ tabella + servizi
+ as.numeric(as.character(prezzo)) | -1, data = palestra_mlogit)
lrtest(m3, m2) #comparison between models
```

```
Likelihood ratio test

Model 1: choice ~ abbonamento + attivita + orario + tabella + servizi +
as.numeric(as.character(prezzo)) | -1
Model 2: choice ~ abbonamento + attivita + orario + tabella + servizi +
prezzo | -1
#Df LogLik Df Chisq Pr(>Chisq)
1 9 -5682.5
2 10 -5677.1 1 10.79 0.001021 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

This comparison leads instead to a p value of 0.001. It means that this new specification significantly reduce the goodness of fit and it is better to keep price variable as qualitative.

Nevertheless 0.001 may be considered a borderline result in terms of statistical significance and choosing the highest significance level, we might consider the model with quantitative price variable and try to compute the willingness-to-pay, since interpreting the measures of part-worths is quite difficult. For example for the personalized schedule level we can compute the following quantity expressed in euros:

Hide

```
coef(m3) ["tabellapersonalizzata"] / (coef(m3) ["as.numeric(as.character(prezzo))"])
```

```
tabellapersonalizzata
25.99332
```

The interpretation is that 26 euros is the price at which customers tend to become indifferent between the two schedule options (standard or personalized).

## Simulate Preference Shares

Together with willingness-to-pay measures, another useful approach to assess the role of subscription attributes consists on using the model to obtain preference share predictions.  
First of all we have to define the “predict.mnl” function.

Hide

```
predict.mnl <- function(model, data) {  
  
  data.model <- model.matrix(update(model$formula, 0 ~ .), data = data)[,-1]  
  logitUtility <- data.model%*%model$coef  
  share <- exp(logitUtility)/sum(exp(logitUtility))  
  cbind(share, data)  
}
```

In order to use “predict.mnl” we have to define a data frame containing the set of designs for which we want to predict the preference shares. To do that, we first create the full set of possible designs:  
And then we choose a reasonable subset where the first row indicates our planned design and the others rows indicate realistic designs offered by competitors:

Hide

```
new_data <- allDesign[c(12, 85, 274,140,185,132), ]  
new_data
```

	abbonamento <fctr>	attività <fctr>	orario <fctr>	tabella <fctr>	servizi <fctr>	prezzo <fctr>
12	annuale	attrezzi+corsi	18_23	standard	consulenza.fisioterapica	40
85	mensile	attrezzi	7_12	standard	non.inclusi	40
274	mensile	attrezzi+corsi	allDay	personalizzata	centro.benessere	80
140	trimestrale	attrezzi	7_12	personalizzata	consulenza.fisioterapica	60
185	trimestrale	attrezzi+corsi	allDay	standard	non.inclusi	60
132	annuale	attrezzi+corsi	allDay	personalizzata	consulenza.fisioterapica	60

6 rows

Now we can predict the preference shares:

Hide

```
predict.mnl(m2, new_data)
```

	share <dbl>	abbonamento <fctr>	attività <fctr>	orario <fctr>	tabella <fctr>	servizi <fctr>	prezzo <fctr>
12	0.892295710	annuale	attrezzi+corsi	18_23	standard	consulenza.fisioterapica	40
85	0.037509982	mensile	attrezzi	7_12	standard	non.inclusi	40
274	0.001994762	mensile	attrezzi+corsi	allDay	personalizzata	centro.benessere	80
140	0.018952457	trimestrale	attrezzi	7_12	personalizzata	consulenza.fisioterapica	60
185	0.007878714	trimestrale	attrezzi+corsi	allDay	standard	non.inclusi	60
132	0.041368375	annuale	attrezzi+corsi	allDay	personalizzata	consulenza.fisioterapica	60

6 rows

Hide

NA

We expect consumers to choose our product design basically 89% of times.  
It useful to get not only the point estimates but also the bootstrap confidence intervals for the preference shares. Indeed predictions are not reliable if the level of accertainty is not declared.

Hide

```
source("BootCI.predict.mnl.R") #load the function  
library(parallel)  
BootCI.predict.mnl(m2 , new_data, nsim = 500, conflevel = 0.95) #use the default values
```

	share <dbl>	2.5% <dbl>	97.5% <dbl>	abbonamento <fctr>	attività <fctr>	orario <fctr>	tabella <fctr>	
12	0.892295710	0.878593094	0.906149668	annuale	attrezzi+corsi	18_23	standard	
85	0.037509982	0.031590287	0.043577480	mensile	attrezzi	7_12	standard	
274	0.001994762	0.001601032	0.002483956	mensile	attrezzi+corsi	allDay	personalizzata	
140	0.018952457	0.015707875	0.022436234	trimestrale	attrezzi	7_12	personalizzata	
185	0.007878714	0.006463266	0.009512343	trimestrale	attrezzi+corsi	allDay	standard	
132	0.041368375	0.035228241	0.047236652	annuale	attrezzi+corsi	allDay	personalizzata	

Hide

NA

## Sensitivity chart

It can be useful to plot the expected changes in preference shares by plotting the sensitivity chart for our planned subscription design with: yearly duration, fitness room plus courses, 18-23 time slot, standard training schedule, physiotherapy consultancy included for a price of 40/month.

Hide

```
sensitivity.mnl <- function(model, attrib, base.data, competitor.data) {

  data <- rbind(base.data, competitor.data)
  base.share <- predict.mnl(model, data)[1,1]
  share <- NULL
  for (a in seq_along(attrib)) {
    for (i in attrib[[a]]) {
      data[1,] <- base.data
      data[1,a] <- i
      share <- c(share, predict.mnl(model, data)[1,1])
    }
  }
  data.frame(level=unlist(attrib), share=share, increase=share-base.share)
}
```

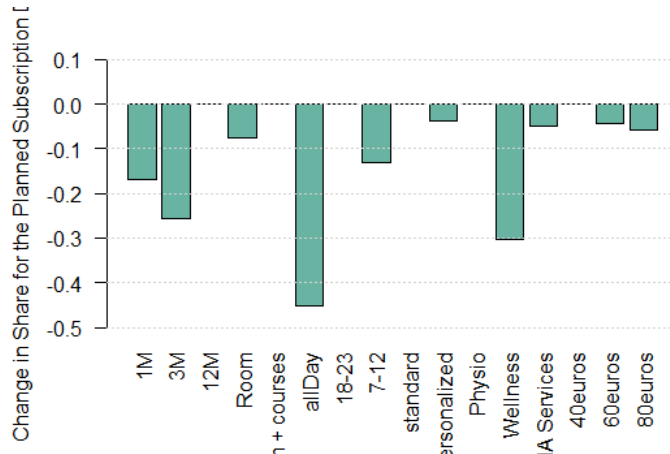
Hide

```
base.data <- new_data[1,]
competitor.data <- new_data[-1,]
(tradeoff <- sensitivity.mnl(m2, attributes, base.data, competitor.data)) #importance of the attributes (preference share)
tradeoff$level
```

Hide

```
barplot(tradeoff$increase, horiz=F, names.arg=c("1M", "3M", "12M", "Room", "Room + courses", "allDay", "18-23", "7-12",
  "standard", "personalized", "Physio", "Wellness", "NA Services", "40euros", "60euros", "80euros"),
  las=2,
  col="#69b3a2",
  ylab="Change in Share for the Planned Subscription Design",

  ylim=c(-0.5,0.1))
grid(nx=NA, ny=NULL)
```



We can see that each change in the attribute levels of the planned design leads to a preference share decrease. In particular the worst changes would be:

- In the time slot attribute: from 18-23 level to All Day level. In this case the preference share decreases by almost 0.5
- In the additional services attribute: from physiotherapy consultancy to access to wellness center. In this case the preference share decreases by 0.3
- In the duration attribute: from yearly duration level to 3-months level. In this case the preference share decreases by almost 0.3

## MIXED MULTINOMIAL LOGIT MODEL

We want now to perform a mixed multinomial logit model in order to deal with consumer heterogeneity.

To estimate it using "mlogit", we have to define a vector called "m2.rpar" indicating which coefficients should vary across customers. The vector must have the same length as the coefficient vector with a letter code indicating what distribution the random coefficients follow across the respondents, for example N for normal. Typically, we assume that all the coefficients are normally distributed across the population.

Hide

```
rpar <- rep("n", length=length(m2$coef))
names(rpar) <- names(m2$coef)
```

We pass this vector to mlogit as the rpar parameter, which stands for random parameters. In addition, we tell mlogit that we have multiple choice observations for each respondent (panel=TRUE) and that we do not want to allow the random parameters to be correlated with each other (correlation=FALSE)

Hide

```
m2_mixed <- mlogit(choice ~ abbonamento +attivita + orario+ tabella + servizi + prezzo| -1, data = palestra_mlogi
t,
                panel=TRUE, rpar = rpar, correlation = FALSE)
```

Hide

summary(m2\_mixed)

```
Call:
mlogit(formula = choice ~ abbonamento + attivita + orario + tabella +
servizi + prezzo | -1, data = palestra_mlogit, rpar = rpar,
correlation = FALSE, panel = TRUE)
```

Frequencies of alternatives:choice

1	2	3	4
0.25633	0.25500	0.24150	0.24717

bfgs method  
21 iterations, 0h:0m:34s  
g'(-H)^-lg = 6.99E-07  
gradient close to zero

Coefficients :			
	Estimate	Std. Error	z-value
abbonamentotrimestrale	-0.486017	0.054078	-8.9874
abbonamentoannuale	1.294231	0.049166	26.3237
attivitaattrezzi+corsi	0.727209	0.040848	17.8028
orario18_23	2.715557	0.065606	41.3919
orario7_12	1.568336	0.058431	26.8407
tabellapersonalizzata	-0.412262	0.040166	-10.2640
servizicentro.benessere	-2.050825	0.061225	-33.4967
servizinon.inclusi	-0.489532	0.045176	-10.8362
prezzo60	-0.491440	0.048587	-10.1147
prezzo80	-0.560255	0.048866	-11.4651
sd.abbonamentotrimestrale	0.542538	0.070624	7.6821
sd.abbonamentoannuale	0.223251	0.072308	3.0875
sd.attivitaattrezzi+corsi	0.326159	0.067601	4.8248
sd.orario18_23	0.798080	0.053542	14.9057
sd.orario7_12	-0.139912	0.076097	-1.8386
sd.tabellapersonalizzata	0.383228	0.062030	6.1781
sd.servizicentro.benessere	0.877216	0.066219	13.2472
sd.servizinon.inclusi	0.540150	0.064597	8.3618
sd.prezzo60	0.693683	0.062119	11.1670
sd.prezzo80	0.048385	0.087589	0.5524
Pr(> z )			
abbonamentotrimestrale	< 2.2e-16 ***		
abbonamentoannuale	< 2.2e-16 ***		
attivitaattrezzi+corsi	< 2.2e-16 ***		
orario18_23	< 2.2e-16 ***		
orario7_12	< 2.2e-16 ***		
tabellapersonalizzata	< 2.2e-16 ***		
servizicentro.benessere	< 2.2e-16 ***		
servizinon.inclusi	< 2.2e-16 ***		
prezzo60	< 2.2e-16 ***		
prezzo80	< 2.2e-16 ***		
sd.abbonamentotrimestrale	1.554e-14 ***		
sd.abbonamentoannuale	0.002018 **		
sd.attivitaattrezzi+corsi	1.402e-06 ***		
sd.orario18_23	< 2.2e-16 ***		
sd.orario7_12	0.065973 .		
sd.tabellapersonalizzata	6.487e-10 ***		
sd.servizicentro.benessere	< 2.2e-16 ***		
sd.servizinon.inclusi	< 2.2e-16 ***		
sd.prezzo60	< 2.2e-16 ***		
sd.prezzo80	0.580670		
---			
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1			

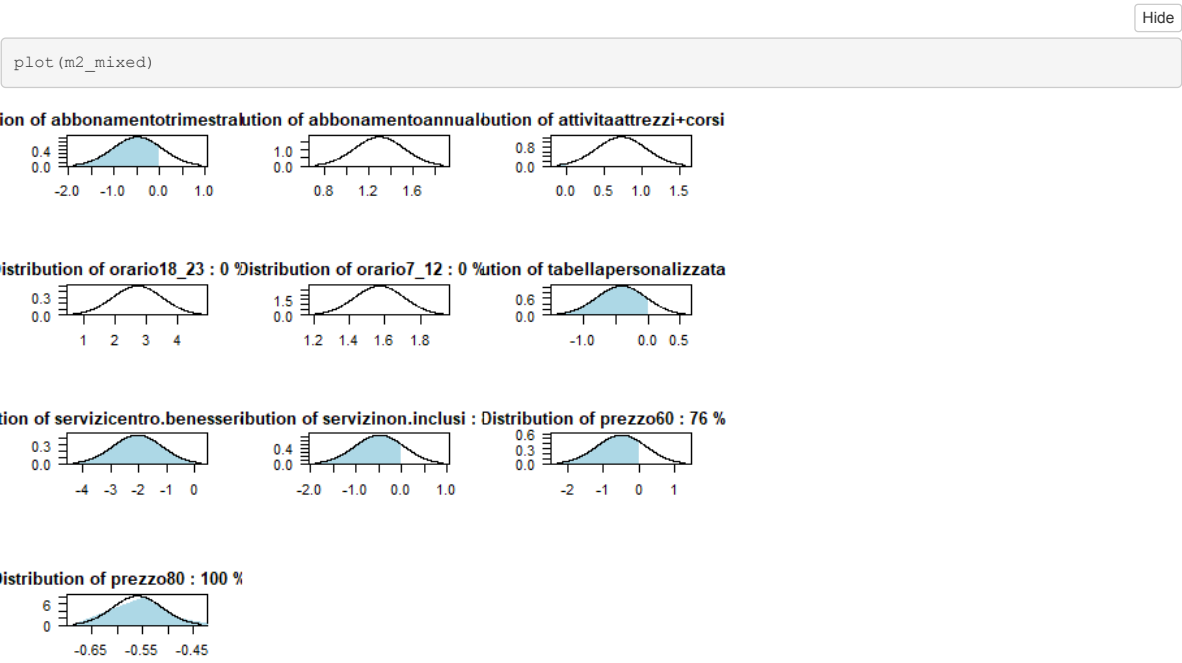
Log-Likelihood: -5459.5

random coefficients				
	Min.	1st Qu.	Median	Mean
abbonamentotrimestrale	-Inf	-0.8519530	-0.4860169	-0.4860169
abbonamentoannuale	-Inf	1.1436509	1.2942314	1.2942314
attivitaattrezzi+corsi	-Inf	0.5072180	0.7272089	0.7272089
orario18_23	-Inf	2.1772597	2.7155565	2.7155565
orario7_12	-Inf	1.4739668	1.5683362	1.5683362
tabellapersonalizzata	-Inf	-0.6707451	-0.4122620	-0.4122620
servizicentro.benessere	-Inf	-2.6424986	-2.0508253	-2.0508253
servizinon.inclusi	-Inf	-0.8538577	-0.4895320	-0.4895320
prezzo60	-Inf	-0.9593217	-0.4914399	-0.4914399
prezzo80	-Inf	-0.5928898	-0.5602548	-0.5602548
3rd Qu. Max.				
abbonamentotrimestrale	-0.12008080	Inf		
abbonamentoannuale	1.44481199	Inf		
attivitaattrezzi+corsi	0.94719982	Inf		
orario18_23	3.25385331	Inf		
orario7_12	1.66270557	Inf		
tabellapersonalizzata	-0.15377895	Inf		
servizicentro.benessere	-1.45915192	Inf		
servizinon.inclusi	-0.12520635	Inf		
prezzo60	-0.02355802	Inf		
prezzo80	-0.52761978	Inf		

For each level attribute we estimate now two parameters instead of one: the estimated average coefficients and the estimated standard deviations, indicating the variability of coefficients across respondents. In general the higher the standard deviation the higher the heterogeneity. In particular we may have to consider the heterogeneity in the preference for:

- the yearly subscription over the three-months one: while consumers on average prefer the annual subscription, there is a non-negligible fraction of them which prefer the three-month one
- the physiotherapy consultancy as an additional service over having additional services not included
- the lower price over the one of 60 euro

The random effects do not show sign reversals in the quartiles distribution but we can check the visual summary of the distribution of the random effects and hence of the level of heterogeneity:



As the plot shows, there is a narrow market niche of customers who has a stronger preference for the three-months subscription with respect to average customers preference and there is also a narrow niche who has a stronger preference for having the additional services not included with respect to other customers.

## Mixed MLN Model with correlated random effects

The model above has covariances among random effects equal to zero, but with correlated random effect coefficients one can asses whether consumers who favor one attribute level also tend to favor another attribute level. So now we use the update function provided by mlogit to estimate the mixed MNL model with correlated random effects:

Hide

```
m2_mixed2 <- update(m2_mixed, correlation = TRUE)
```

We can test the hypothesis that the random parameters are uncorrelated:

Hide

```
waldtest(m2_mixed2, correlation = FALSE)
```

Wald test  
  
data: uncorrelated random effects  
chisq = 428.44, df = 45, p-value < 2.2e-16

The test clearly reject the hypothesis that the random parameters are uncorrelated.  
To get a better sense of the strength of the association among random coefficients, we extract the covariance matrix:

Hide

```
as.data.frame(cov2cor(cov.mlogit(m2_mixed2)))
```

	abbonamentotrimestrale <dbl>	abbonamentoannuale <dbl>	attivitaattrezzi+corsi <dbl>
abbonamentotrimestrale	1.00000000	-0.85069212	0.10521163
abbonamentoannuale	-0.85069212	1.00000000	0.06319491



attivitaaattrezzi+corsi	0.10521163	0.06319491	1.00000000
orario18_23	0.07520412	-0.22695057	0.36497683
orario7_12	0.28574675	-0.28025421	0.47121571
tabellapersonalizzata	0.08982321	0.39590559	0.18106125
servizicentro.benessere	0.05561454	-0.04087804	-0.13085697
servizinon.inclusi	0.08506212	-0.02408728	-0.48256262
prezzo60	-0.42532620	0.43108644	0.13685190
prezzo80	-0.52023335	0.33933200	-0.44471712

1-10 of 10 rows | 1-4 of 10 columns

Hide

NA

The covariance matrix shows strong associations between the part worth for the following attributes and levels:

- a negative association between yearly subscriptions and the three-months one
- a positive association between 18-23 time slot and 7-12 time slot
- a positive association between wellness center and additional service non included

. We obtain the standard errors of the correlations among random effects and hence perform significance test in order to see which correlations are significant:

Hide

```
summary(vcov(m2_mixed2, what = "rpar", type = "cor"))
```

	Estimate
sd.abbonamentotrimestrale	0.5206406
sd.abbonamentoannuale	0.1193403
sd.attivitaaattrezzi+corsi	0.3199233
sd.orario18_23	1.2757584
sd.orario7_12	0.5467932
sd.tabellapersonalizzata	0.4243548
sd.servizicentro.benessere	1.1898518
sd.servizinon.inclusi	0.7762340
sd.prezzo60	0.8041677
sd.prezzo80	0.2097898
cor.abbonamentotrimestrale:abbonamentoannuale	-0.8506921
cor.abbonamentotrimestrale:attivitaaattrezzi+corsi	0.1052116
cor.abbonamentoannuale:attivitaaattrezzi+corsi	0.0631949
cor.abbonamentotrimestrale:orario18_23	0.0752041
cor.abbonamentoannuale:orario18_23	-0.2269506
cor.attivitaaattrezzi+corsi:orario18_23	0.3649768
cor.abbonamentotrimestrale:orario7_12	0.2857467
cor.abbonamentoannuale:orario7_12	-0.2802542
cor.attivitaaattrezzi+corsi:orario7_12	0.4712157
cor.orario18_23:orario7_12	0.9271161
cor.abbonamentotrimestrale:tabellapersonalizzata	0.0898232
cor.abbonamentoannuale:tabellapersonalizzata	0.3959056
cor.attivitaaattrezzi+corsi:tabellapersonalizzata	0.1810612
cor.orario18_23:tabellapersonalizzata	-0.1500685
cor.orario7_12:tabellapersonalizzata	0.1353971
cor.abbonamentotrimestrale:servizicentro.benessere	0.0556145
cor.abbonamentoannuale:servizicentro.benessere	-0.0408780
cor.attivitaaattrezzi+corsi:servizicentro.benessere	-0.1308570
cor.orario18_23:servizicentro.benessere	0.0801337
cor.orario7_12:servizicentro.benessere	0.0884963
cor.tabellapersonalizzata:servizicentro.benessere	0.1620889
cor.abbonamentotrimestrale:servizinon.inclusi	0.0850621
cor.abbonamentoannuale:servizinon.inclusi	-0.0240873
cor.attivitaaattrezzi+corsi:servizinon.inclusi	-0.4825626
cor.orario18_23:servizinon.inclusi	0.0562673
cor.orario7_12:servizinon.inclusi	0.0048703
cor.tabellapersonalizzata:servizinon.inclusi	0.1151529
cor.servizicentro.benessere:servizinon.inclusi	0.6752330
cor.abbonamentotrimestrale:prezzo60	-0.4253262
cor.abbonamentoannuale:prezzo60	0.4310864
cor.attivitaaattrezzi+corsi:prezzo60	0.1368519
cor.orario18_23:prezzo60	-0.1119757
cor.orario7_12:prezzo60	-0.2944581
cor.tabellapersonalizzata:prezzo60	-0.0514435
cor.servizicentro.benessere:prezzo60	-0.1567349
cor.servizinon.inclusi:prezzo60	-0.1434078
cor.abbonamentotrimestrale:prezzo80	-0.5202333
cor.abbonamentoannuale:prezzo80	0.3393320
cor.attivitaaattrezzi+corsi:prezzo80	-0.4447171
cor.orario18_23:prezzo80	-0.1785393
cor.orario7_12:prezzo80	-0.3803329
cor.tabellapersonalizzata:prezzo80	-0.2350627
cor.servizicentro.benessere:prezzo80	-0.3777922
cor.servizinon.inclusi:prezzo80	-0.1788344
cor.prezzo60:prezzo80	0.4746354
Std. Error	
sd.abbonamentotrimestrale	0.0788431
sd.abbonamentoannuale	0.0662171
sd.attivitaaattrezzi+corsi	0.0573346
sd.orario18_23	0.0796460
sd.orario7_12	0.0727785
sd.tabellapersonalizzata	0.0553035

sd.servizicentro.benessere	0.0750898
sd.servizinon.inclusi	0.0637763
sd.prezzo60	0.0696691
sd.prezzo80	0.0770579
cor.abbonamentotrimestrale:abbonamentoannuale	0.7722450
cor.abbonamentotrimestrale:attivitaaattrezzi+corsi	0.3136430
cor.abbonamentoannuale:attivitaaattrezzi+corsi	0.5518543
cor.abbonamentotrimestrale:orario18_23	0.1807927
cor.abbonamentoannuale:orario18_23	0.6909648
cor.attivitaaattrezzi+corsi:orario18_23	0.2402033
cor.abbonamentotrimestrale:orario7_12	0.2115763
cor.abbonamentoannuale:orario7_12	0.7151039
cor.attivitaaattrezzi+corsi:orario7_12	0.2092767
cor.orario18_23:orario7_12	0.0728841
cor.abbonamentotrimestrale:tabellapersonalizzata	0.2424595
cor.abbonamentoannuale:tabellapersonalizzata	0.6082889
cor.attivitaaattrezzi+corsi:tabellapersonalizzata	0.2783307
cor.orario18_23:tabellapersonalizzata	0.2539746
cor.orario7_12:tabellapersonalizzata	0.2491982
cor.abbonamentotrimestrale:servizicentro.benessere	0.1664260
cor.abbonamentoannuale:servizicentro.benessere	0.6541863
cor.attivitaaattrezzi+corsi:servizicentro.benessere	0.2745262
cor.orario18_23:servizicentro.benessere	0.1148550
cor.orario7_12:servizicentro.benessere	0.1663826
cor.tabellapersonalizzata:servizicentro.benessere	0.1972877
cor.abbonamentotrimestrale:servizinon.inclusi	0.1701227
cor.abbonamentoannuale:servizinon.inclusi	0.5732913
cor.attivitaaattrezzi+corsi:servizinon.inclusi	0.3155717
cor.orario18_23:servizinon.inclusi	0.1483228
cor.orario7_12:servizinon.inclusi	0.2128334
cor.tabellapersonalizzata:servizinon.inclusi	0.2138794
cor.servizicentro.benessere:servizinon.inclusi	0.0740991
cor.abbonamentotrimestrale:prezzo60	0.2600541
cor.abbonamentoannuale:prezzo60	0.5750003
cor.attivitaaattrezzi+corsi:prezzo60	0.2569550
cor.orario18_23:prezzo60	0.1554845
cor.orario7_12:prezzo60	0.2363580
cor.tabellapersonalizzata:prezzo60	0.2332039
cor.servizicentro.benessere:prezzo60	0.1635154
cor.servizinon.inclusi:prezzo60	0.1802785
cor.abbonamentotrimestrale:prezzo80	0.6223239
cor.abbonamentoannuale:prezzo80	0.6860409
cor.attivitaaattrezzi+corsi:prezzo80	0.6247259
cor.orario18_23:prezzo80	0.5302575
cor.orario7_12:prezzo80	0.5726641
cor.tabellapersonalizzata:prezzo80	0.5155309
cor.servizicentro.benessere:prezzo80	0.5106872
cor.servizinon.inclusi:prezzo80	0.4840012
cor.prezzo60:prezzo80	0.3029266
z-value	
sd.abbonamentotrimestrale	6.6035
sd.abbonamentoannuale	1.8023
sd.attivitaaattrezzi+corsi	5.5799
sd.orario18_23	16.0179
sd.orario7_12	7.5131
sd.tabellapersonalizzata	7.6732
sd.servizicentro.benessere	15.8457
sd.servizinon.inclusi	12.1712
sd.prezzo60	11.5427
sd.prezzo80	2.7225
cor.abbonamentotrimestrale:abbonamentoannuale	-1.1016
cor.abbonamentotrimestrale:attivitaaattrezzi+corsi	0.3355
cor.abbonamentoannuale:attivitaaattrezzi+corsi	0.1145
cor.abbonamentotrimestrale:orario18_23	0.4160
cor.abbonamentoannuale:orario18_23	-0.3285
cor.attivitaaattrezzi+corsi:orario18_23	1.5194
cor.abbonamentotrimestrale:orario7_12	1.3506
cor.abbonamentoannuale:orario7_12	-0.3919
cor.attivitaaattrezzi+corsi:orario7_12	2.2516
cor.orario18_23:orario7_12	12.7204
cor.abbonamentotrimestrale:tabellapersonalizzata	0.3705
cor.abbonamentoannuale:tabellapersonalizzata	0.6509
cor.attivitaaattrezzi+corsi:tabellapersonalizzata	0.6505
cor.orario18_23:tabellapersonalizzata	-0.5909
cor.orario7_12:tabellapersonalizzata	0.5433
cor.abbonamentotrimestrale:servizicentro.benessere	0.3342
cor.abbonamentoannuale:servizicentro.benessere	-0.0625
cor.attivitaaattrezzi+corsi:servizicentro.benessere	-0.4767
cor.orario18_23:servizicentro.benessere	0.6977
cor.orario7_12:servizicentro.benessere	0.5319
cor.tabellapersonalizzata:servizicentro.benessere	0.8216
cor.abbonamentotrimestrale:servizinon.inclusi	0.5000
cor.abbonamentoannuale:servizinon.inclusi	-0.0420
cor.attivitaaattrezzi+corsi:servizinon.inclusi	-1.5292
cor.orario18_23:servizinon.inclusi	0.3794
cor.orario7_12:servizinon.inclusi	0.0229
cor.tabellapersonalizzata:servizinon.inclusi	0.5384
cor.servizicentro.benessere:servizinon.inclusi	9.1126
cor.abbonamentotrimestrale:prezzo60	-1.6355
cor.abbonamentoannuale:prezzo60	0.7497
cor.attivitaaattrezzi+corsi:prezzo60	0.5326
cor.orario18_23:prezzo60	-0.7202
cor.orario7_12:prezzo60	-1.2458
cor.tabellapersonalizzata:prezzo60	-0.2206
cor.servizicentro.benessere:prezzo60	-0.9585
cor.servizinon.inclusi:prezzo60	-0.7955

```

cor.abbonamentotrimetrale:prezzo80 -0.8360
cor.abbonamentoannuale:prezzo80 0.4946
cor.attivitaattrezzi+corsi:prezzo80 -0.7119
cor.orario18_23:prezzo80 -0.3367
cor.orario7_12:prezzo80 -0.6641
cor.tabellapersonalizzata:prezzo80 -0.4560
cor.servizicentro.benessere:prezzo80 -0.7398
cor.servizinon.inclusi:prezzo80 -0.3695
cor.prezzo60:prezzo80 1.5668
Pr(>|z|)

sd.abbonamentotrimetrale 4.015e-11 ***
sd.abbonamentoannuale 0.071505 .
sd.attivitaattrezzi+corsi 2.406e-08 ***
sd.orario18_23 < 2.2e-16 ***
sd.orario7_12 5.774e-14 ***
sd.tabellapersonalizzata 1.678e-14 ***
sd.servizicentro.benessere < 2.2e-16 ***
sd.servizinon.inclusi < 2.2e-16 ***
sd.prezzo60 < 2.2e-16 ***
sd.prezzo80 0.006479 **
cor.abbonamentotrimetrale:abbonamentoannuale 0.270643
cor.abbonamentotrimetrale:attivitaattrezzi+corsi 0.737285
cor.abbonamentoannuale:attivitaattrezzi+corsi 0.908831
cor.abbonamentotrimetrale:orario18_23 0.677433
cor.abbonamentoannuale:orario18_23 0.742568
cor.attivitaattrezzi+corsi:orario18_23 0.128649
cor.abbonamentotrimetrale:orario7_12 0.176836
cor.abbonamentoannuale:orario7_12 0.695127
cor.attivitaattrezzi+corsi:orario7_12 0.024345 *
cor.orario18_23:orario7_12 < 2.2e-16 ***
cor.abbonamentotrimetrale:tabellapersonalizzata 0.711035
cor.abbonamentoannuale:tabellapersonalizzata 0.515143
cor.attivitaattrezzi+corsi:tabellapersonalizzata 0.515353
cor.orario18_23:tabellapersonalizzata 0.554601
cor.orario7_12:tabellapersonalizzata 0.586902
cor.abbonamentotrimetrale:servizicentro.benessere 0.738251
cor.abbonamentoannuale:servizicentro.benessere 0.950175
cor.attivitaattrezzi+corsi:servizicentro.benessere 0.633601
cor.orario18_23:servizicentro.benessere 0.485368
cor.orario7_12:servizicentro.benessere 0.594806
cor.tabellapersonalizzata:servizicentro.benessere 0.411312
cor.abbonamentotrimetrale:servizinon.inclusi 0.617072
cor.abbonamentoannuale:servizinon.inclusi 0.966486
cor.attivitaattrezzi+corsi:servizinon.inclusi 0.126222
cor.orario18_23:servizinon.inclusi 0.704423
cor.orario7_12:servizinon.inclusi 0.981743
cor.tabellapersonalizzata:servizinon.inclusi 0.590300
cor.servizicentro.benessere:servizinon.inclusi < 2.2e-16 ***
cor.abbonamentotrimetrale:prezzo60 0.101938
cor.abbonamentoannuale:prezzo60 0.453426
cor.attivitaattrezzi+corsi:prezzo60 0.594317
cor.orario18_23:prezzo60 0.471418
cor.orario7_12:prezzo60 0.212833
cor.tabellapersonalizzata:prezzo60 0.825408
cor.servizicentro.benessere:prezzo60 0.337794
cor.servizinon.inclusi:prezzo60 0.426335
cor.abbonamentotrimetrale:prezzo80 0.403182
cor.abbonamentoannuale:prezzo80 0.620866
cor.attivitaattrezzi+corsi:prezzo80 0.476552
cor.orario18_23:prezzo80 0.736341
cor.orario7_12:prezzo80 0.506597
cor.tabellapersonalizzata:prezzo80 0.648417
cor.servizicentro.benessere:prezzo80 0.459438
cor.servizinon.inclusi:prezzo80 0.711761
cor.prezzo60:prezzo80 0.117154
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

That confirms the correlations mentioned above. In particular the correlation between `orario18_23:orario7_12` and between `servizicentro.benessere:servizinon.inclusi` are highly significant, showing that respondents have behaved rationally. Indeed it has perfectly sense that a consumer always prefers to go to the gym at a certain time rather than all day.

We can reduce the model complexity by only including the highest significant correlations

Hide

```

m2_mixed3 <- update(m2_mixed2, correlation = c("orario18_23", "orario7_12", "servizicentro.benessere", "servizinon.inclusi" ))

```

The significant presence of random coefficients and their correlation can be further investigated using the ML ratio test:

Hide

```

lrtest(m2, m2_mixed) #Fixed effects vs. uncorrelated random effects

```

Likelihood ratio test

```

Model 1: choice ~ abbonamento + attivita + orario + tabella + servizi +
  prezzo | -1
Model 2: choice ~ abbonamento + attivita + orario + tabella + servizi +
  prezzo | -1
#Df LogLik Df Chisq Pr(>Chisq)
1 10 -5677.1
2 20 -5459.5 10 435.23 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

We can reject the null hypothesis that the variances of the random effects are jointly equal to zero. Therefore we should use the model with random effects.

Hide

```
lrtest(m2_mixed, m2_mixed2) #Uncorrelated random effects vs. all correlated random effects
```

```
Likelihood ratio test

Model 1: choice ~ abbonamento + attivita + orario + tabella + servizi +
  prezzo | -1
Model 2: choice ~ abbonamento + attivita + orario + tabella + servizi +
  prezzo | -1
#Df  LogLik Df  Chisq Pr(>Chisq)
1   20 -5459.5
2   65 -5362.5 45 194.12  < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Moreover we can not assume that random effects are independent, the model with correlated random effects has a better goodness of fit.

Hide

```
lrtest(m2_mixed3, m2_mixed2) #partially correlated random effects vs. all correlated random effects
```

```
Likelihood ratio test

Model 1: choice ~ abbonamento + attivita + orario + tabella + servizi +
  prezzo | -1
Model 2: choice ~ abbonamento + attivita + orario + tabella + servizi +
  prezzo | -1
#Df  LogLik Df  Chisq Pr(>Chisq)
1   26 -5425.5
2   65 -5362.5 39 126.07  4.173e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

And at the end our best model is the one that includes all random effects. Once we have established it, we can compute the preference shares.

## Simulating shares

To compute share predictions with a mixed MNL model, we have to compute the preference shares for each of newly sampled, representative respondents. The part worths for these respondents are drawn from a multivariate normal distribution. Once we have the preference shares for all of the representative respondents, we average across respondents to get our overall preference share predictions. Firstly we have to implement the function for predicting shares from a mixed MNL model

Hide

```
predict.mixed.mnl <- function(model, data, nresp=1000){
  data.model <- model.matrix(update(model$formula, 0 ~ .), data = data)[,-1]
  coef.Sigma <- cov.mlogit(model)
  coef.mu <- model$coef[1:dim(coef.Sigma)[1]]
  draws <- mvrnorm(n=nresp, coef.mu, coef.Sigma)
  shares <- matrix(NA, nrow=nresp, ncol=nrow(data))
  for (i in 1:nresp) {
    utility <- data.model%*%draws[i,]
    share = exp(utility)/sum(exp(utility))
    shares[i,] <- share
  }
  cbind(colMeans(shares), data)
}
```

Now we can compute the preference shares:

Hide

```
set.seed(123)
predict.mixed.mnl(m2_mixed2, data=new_data)
```

	colMeans(shares)	abbonamento	attivita	orario	tabella	servizi	
	<dbl>	<fctr>	<fctr>	<fctr>	<fctr>	<fctr>	
12	0.883788099	annuale	attrezzi+corsi	18_23	standard	consulenza.fisioterapica	
85	0.038046571	mensile	attrezzi	7_12	standard	non.inclusi	
274	0.002178193	mensile	attrezzi+corsi	allDay	personalizzata	centro.benessere	
140	0.015424742	trimestrale	attrezzi	7_12	personalizzata	consulenza.fisioterapica	
185	0.007829564	trimestrale	attrezzi+corsi	allDay	standard	non.inclusi	
132	0.052732831	annuale	attrezzi+corsi	allDay	personalizzata	consulenza.fisioterapica	

6 rows | 1-7 of 7 columns

As we can see, preference shares of the random effects model are slightly different with respect to those seen in the fixed effects model. That happens since the mixed model better account for the costumer heterogeneity, by predicting larger preference shares to niche product. For the sake of completeness we point out that also in this case it is possible to calculate the confidence intervals of the preference shares, even if we do not implement the code due to its high complexity

Hide

```
source("BootCI.predict.mixed.mnl.R") #load the function
library(parallel)
```

```
BootCI.predict.mixed.mnl(m2_mixed2, new_data, nsim = 500, conflevel = 0.95, nresp=1000) #use the default values
```

## Conclusions

We have found out that our best model is the mixed multinomial logit model that includes all random effects. According to consumers preferences, gym subscriptions should have these characteristics: of yearly duration, including fitness room plus courses, allowing the access in the 18-23 time slot, including the standard training schedule , including a physiotherapy consultancy as an additional service, price of 40 euros/month. Thank to this kind of model, we also have been able to control for the circumstance that there are small fractions of respondents for which "niche products", such as three-months subscription and additional services not included, have great value.