Creation_data

Marta Bagno

05 settembre 2018

```
# Load the libarary
library("dplyr")
```

```
# 6 choice sets for 484 individuals, with 4 choice options each.
The fourth alternative is "none of these".
R= 484
n_choice_set=6
N= R*n_choice_set
S= 4
A=n_choice_set*S
P= 11
N2= N*S
altern=rep(1:S, time=N)
length(altern)
```

```
## [1] 11616
```

```
table(altern)
```

```
## altern
##    1    2    3    4
## 2904 2904 2904 2904
```

```
# The function "indexes" used to obtain the first and the last observation in each
 choice set is taken from the
# link: http://rpubs.com/jimsavage/using_decisionmaker_variables_in_mixed_logit
indexes <- data_frame(task = rep(1:N, each = S),
                      row = 1:(N*S)) %>%
  group_by(task) %>%
  summarise(start = first(row),
            end = last(row))

choice_c=c(rep(1:n_choice_set, each = S, times=R))
length(choice_c)
```

```
## [1] 11616
```

```
# index for respondent
index= c(rep(1:R, each = A))
table(index)
```

```
## index
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
```

```
##  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
##  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
##  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
##  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
##  91  92  93  94  95  96  97  98  99 100 101 102 103 104 105 106 107 108
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
## 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484
##  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24  24
```

```
# choice made by respondent
choice=c(rep(NA,times=N2))
for(i in 1:N){
a=c(rep(0,times=S))
b=sample(1:S,1)
a[b]=1
choice[indexes$start[i]:indexes$end[i]]=a
}
table(choice)
```

```
## choice
##    0    1
## 8712 2904
```

```
# dummy for each alternative except for the fourth
alt1=c(rep(NA,times=N2))
for (i in 1:N2){
if (altern[i]==1){
alt1[i]=1
}else {alt1[i]=0}
}
table(alt1)
```

```
## alt1
##    0    1
## 8712 2904
```

```
alt2=c(rep(NA,times=N2))
for (i in 1:N2){
  if (altern[i]==2){
    alt2[i]=1
  }else {alt2[i]=0}
}
table(alt2)
```

```
## alt2
##    0    1
## 8712 2904
```

```
alt3=c(rep(NA,times=N2))
for (i in 1:N2){
  if (altern[i]==3){
    alt3[i]=1
  }else {alt3[i]=0}
}
table(alt3)
```

```
## alt3
##    0    1
## 8712 2904
```

```r
# matrix of some predictor
X <- matrix(c(alt1, alt2, alt3, rnorm(N2*8)), N2, 11)
# replace the values of fourth row (the row of "none of these") with zero
for (i in 1:N){
X[i*S, ] = 0
}

dd <- data.frame(choice_c, index, choice,altern, X )
class(dd)
```

```
## [1] "data.frame"
```

```r
colnames(dd)
```

```
##  [1] "choice_c" "index"    "choice"   "altern"   "X1"       "X2"
##  [7] "X3"       "X4"       "X5"       "X6"       "X7"       "X8"
## [13] "X9"       "X10"      "X11"
```

```r
colnames(dd)[c(5:15)]  <- c("alt1","alt2","alt3", "X1","X2", "X3", "X4","X5", "X6", "X7","X8")
colnames(dd)
```

```
##  [1] "choice_c" "index"    "choice"   "altern"   "alt1"     "alt2"
##  [7] "alt3"     "X1"       "X2"       "X3"       "X4"       "X5"
## [13] "X6"       "X7"       "X8"
```

```r
head(dd)
```

```
##   choice_c index choice altern alt1 alt2 alt3          X1         X2
## 1        1     1      0      1    1    0    0  0.18189538 -1.5505448
## 2        1     1      0      2    0    1    0 -1.39443282 -1.8517668
## 3        1     1      0      3    0    0    1 -0.06147667 -0.7467376
## 4        1     1      1      4    0    0    0  0.00000000  0.0000000
## 5        2     1      1      1    1    0    0 -1.76015182  0.2981056
## 6        2     1      0      2    0    1    0  1.23091515  2.5947576
##           X3         X4         X5         X6         X7         X8
## 1 -1.0635264  1.1794433  0.6997751  0.2134426 -0.7283559 -0.3125102
## 2 -0.6549009  1.2148332 -1.1981096  0.7917094  0.2126527  1.5367807
## 3 -1.1732590 -1.5668760  0.1198097 -0.1377535  0.3074851 -2.4811549
## 4  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000  0.0000000
## 5  0.2292465  1.2822815 -0.3850197 -0.3734933 -1.0137362 -0.9993479
## 6 -0.6357539 -0.5825819 -0.7939223  0.4505146  0.2483617 -0.5184921
```

```r
# create a useful y variable for the three models
y= dd[dd$choice== 1, "altern"]
table(y)
```

```
## y
##   1   2   3   4
## 717 702 741 744
```

```r
# extract the respondent index for the alternative choice
id = dd[dd$choice == 1, "index"]
table(id)
```

```
## id
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
##  19  20  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
##  37  38  39  40  41  42  43  44  45  46  47  48  49  50  51  52  53  54
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
##  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
##  73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
##  91  92  93  94  95  96  97  98  99 100 101 102 103 104 105 106 107 108
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
```

```
## 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
## 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484
##   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6   6
```