Econ 613 - Assignment 3

Zixuan Qi

March 24, 2022

Exercise 1

Question 1

The number of students is 340823; the number of schools is 898; the number of programs is 32.

Question 2

The number of choices is 2773.

Question 3

The number of students applying to at least one senior high schools in the same district to home is 265464.

Question 4

The number of students each senior high school admitted is as follows.

```
# A tibble: 6 x 2
  schoolcode num_stu
       <int>   <int>
1      10101     398
2      10102     248
3      10103     443
4      10104     220
5      10105     346
6      10106     395
```

Figure 1: Number of Admitted Students

Question 5

The cutoff of senior high schools is as follows.

```
# A tibble: 6 x 2
  schoolcode min_score
       <int>     <int>
1      10101       284
2      10102       343
3      10103       316
4      10104       245
5      10105       260
6      10106       293
```

Figure 2: Cutoff

Question 6

The quality of senior high schools is as follows.

```
# A tibble: 6 x 2
  schoolcode avg_score
       <int>     <dbl>
1      10101      320.
2      10102      394.
3      10103      354.
4      10104      297.
5      10105      351.
6      10106      340.
```

Figure 3: Quality

Exercise 2

The school level dataset is as follows.

```
# A tibble: 6 x 8
  choice       rankplace  size cutoff quality sssdistrict ssslong ssslat
  <chr>            <dbl> <int>  <int>   <dbl> <chr>         <dbl>  <dbl>
1 100101,Gene~         3    79    198    244. Wa Municip~   -2.29   10.0
2 100101,Home~         3    40    199    229. Wa Municip~   -2.29   10.0
3 100101,Tech~         4    49    201    235. Wa Municip~   -2.29   10.0
4 100102,Agri~         1    90    273    293. Wa Municip~   -2.29   10.0
5 100102,Busi~         2    90    283    303. Wa Municip~   -2.29   10.0
6 100102,Gene~         3    90    291    311. Wa Municip~   -2.29   10.0
```

Figure 4: School Level Dataset

Exercise 3

The distance is as follows.

```
  V1    schoolcode Programs  schoolname     sssdistrict   ssslong ssslat jssdistrict   jsslong jsslat distance
  <chr>      <int> <chr>     <chr>          <chr>           <dbl>  <dbl> <chr>           <dbl>  <dbl>    <dbl>
1 1          50112 Home Eco~ KUMASI SENIOR~ Kumasi Metro    -1.60   6.68 Bosomtwe/Atw~   -1.56   6.56     8.81
2 1          50107 General ~ ANGLICAN SENI~ Kumasi Metro    -1.60   6.68 Bosomtwe/Atw~   -1.56   6.56     8.81
3 1          50202 Visual A~ TOASE SENIOR ~ Atwima / Nwa~   -1.81   6.68 Bosomtwe/Atw~   -1.56   6.56    18.9
4 1          50202 Visual A~ TOASE SENIOR ~ Atwima / Nwa~   -1.81   6.68 Bosomtwe/Atw~   -1.56   6.56    18.9
5 1          50702 Home Eco~ SIMMS SENIOR ~ Kwabre (Mamp~   -1.54   6.81 Bosomtwe/Atw~   -1.56   6.56    17.2
6 1          50901 General ~ EJURAMAN ANGL~ Ejura/Sekyed~   -1.37   7.46 Bosomtwe/Atw~   -1.56   6.56    63.9
```

Figure 5: Distance

Exercise 4

Question 1

Recode the schoolcode into its first three digits.

stu_dataset.schoolcode	stu_dataset.scode_rev
50112	501
50107	501
50202	502
50202	502
50702	507
50901	509

Figure 6: Revised Code

Question 2

Recode the program variable into 4 categories.

stu_dataset.Programs	stu_dataset.pgm_rev
Home Economics	Economics
General Arts	Arts
Visual Arts	Arts
Visual Arts	Arts
Home Economics	Economics
General Arts	Arts

Figure 7: Revised Program

Question 3

Create a new choice variable choice_rev.

stu_dataset.scode_rev	stu_dataset.pgm_rev	stu_dataset.choice_rev
501	Economics	501 Economics
501	Arts	501 Arts
502	Arts	502 Arts
502	Arts	502 Arts
507	Economics	507 Economics
509	Arts	509 Arts

Figure 8: Revised Choice

Question 4

Recalculate the cutoff and the quality for each recoded choice.

```
# A tibble: 6 x 3
  choice_rev    rev_cutoff rev_quality
  <chr>              <int>       <dbl>
1 304 Arts             207        295.
2 304 Others           219        319.
3 304 Economics        192        298.
4 210 Science          206        333.
5 210 Arts             208        291.
6 210 Economics        203        294.
```

Figure 9: Revised Cutoff and Quality

Question 5 & 6

Consider the 20,000 highest score students. To facilitate the subsequent calculation, I only retain each students' first choice in this question.

```
# A tibble: 6 x 4
  V1     choice_rev  rev_quality score
  <chr>  <chr>             <dbl> <int>
1 335624 301 Science        411.   469
2 318458 210 Science        333.   468
3 318492 210 Science        333.   467
4 335584 301 Science        411.   467
5 318422 210 Science        333.   466
6 318525 210 Science        333.   466
```

Figure 10: 20,000 Highest Score Students

Exercise 5

Question 1

To save the calculation time, I only regress the first choice on intercepts and the test score. The likelihood function can be seen in the mllike function.

Question 2

The number of coefficients should be $2 * 245$. The parameters can be seen in the results of *model1*. The number of marginal effects should be 246. The values can be seen in the results of *me_model1*.

```
> dim(model1$par)
[1]   2 245
```

(a) Coefficients

```
> dim(as.matrix(me_model1))
[1] 246   1
```

(b) Marginal Effects

Figure 11: Dimension of First Model

Exercise 6

Question 1

To save the calculation time, I only regress the first choice on intercepts and the school quality. The likelihood function can be seen in mclike function.

Question 2

The number of coefficients should be 246. The parameters can be seen in the results of *model2*. The number of marginal effects should be $246 * 246$. The values can be seen in the results of *me_model2*.

```
> dim(as.matrix(model2$par))
[1] 246   1
```

(a) Coefficients

```
> dim(me_model2)
[1] 246 246
```

(b) Marginal Effects

Figure 12: Dimension of Second Model

Exercise 7

Question 1

In my opinion, the second model is appropriate in this exercise. The reason is that in the first model, the variable "score" is alternative-invariant, which means this variable does not vary with alternatives. Thus, if I use the first model and exclude alternatives containing "Others", the results may not change. In contrast, in the second model, the variable "quality" is alternative-variant, meaning that this variable varies with alternatives. Thus, if I exclude alternatives containing "Others", the results will change.

Question 2 & 3

The appropriate model is the second model, thus before excluding the alternatives containing "Others", the results are same as Exercise 6. They can be seen in *pj*. The number of probabilities should be 246. The probabilities after excluding the alternatives containing "Others" can be seen in *pj2*. The number of probabilities should be 196.

```
> dim(as.matrix(pj2))
[1] 196   1
> dim(as.matrix(pj))
[1] 246   1
```

Figure 13: Probabilities