Assignment Three Outputs

Exercise 1

1. For the first question:

   (a) The number of students is 340823.

   (b) Judging by the school code, the number of school is 898. Judging by the school name, the number of school is 651.

   (c) The number of program is 32 (there's a NA value).

2. The number of choice is 2773 after removing NA values of the school choice and the program choice.

3. The number of students applying to at least one senior high schools in the same district to home is 262085.

4. The number of students of each school admitted is provided in the "Nadmit" table. Here is the head of it with the first 10 schools:

	schoolcode	Nadmit
1	10101	398
2	10102	248
3	10103	443
4	10104	220
5	10105	346
6	10106	395
7	10107	306
8	10108	318
9	10109	300
10	10110	535

5. The cutoff score of each school is provided in the "lowscore" table. Here is the head of it with the first 10 schools:

	schoolcode	minscore
1	10101	284
2	10102	343
3	10103	316
4	10104	245
5	10105	260
6	10106	293
7	10107	281
8	10108	248
9	10109	257
10	10110	343

6. The average score of each school is provided in the "quality" table. Here is the head of it with the first 10 schools:

	schoolcode	meanscore
1	10101	320.2312
2	10102	394.1492
3	10103	353.8330
4	10104	296.9182
5	10105	351.2139
6	10106	340.1013
7	10107	311.9542
8	10108	303.9025
9	10109	281.8233
10	10110	408.0785

Exercise 2

1. The "schooldata" contains all required information.

2. Here is the head of the dataset with first 10 observations.

	choice	schoolcode	schoolname	sssdistrict	ssslong	ssslat	pchoice	people	minscore	meanscore
1	100101 General Arts	100101	WA SENIOR HIGH/TECHNICAL SCHOOL, WA	Wa Municipal	-2.28503036	10.030622	General Arts	79	198	244.3924
2	100101 Home Economics	100101	WA SENIOR HIGH/TECHNICAL SCHOOL, WA	Wa Municipal	-2.28503036	10.030622	Home Economics	40	199	229.4500
3	100101 Technical	100101	WA SENIOR HIGH/TECHNICAL SCHOOL, WA	Wa Municipal	-2.28503036	10.030622	Technical	49	201	235.1020
4	100102 Agriculture	100102	WA SENIOR HIGH SCHOOL WA	Wa Municipal	-2.28503036	10.030622	Agriculture	90	273	292.5556
5	100102 Business	100102	WA SENIOR HIGH SCHOOL WA	Wa Municipal	-2.28503036	10.030622	Business	90	283	303.3444
6	100102 General Arts	100102	WA SENIOR HIGH SCHOOL WA	Wa Municipal	-2.28503036	10.030622	General Arts	90	291	311.1111
7	100102 General Science	100102	WA SENIOR HIGH SCHOOL WA	Wa Municipal	-2.28503036	10.030622	General Science	90	273	298.4333
8	100102 Home Economics	100102	WA SENIOR HIGH SCHOOL WA	Wa Municipal	-2.28503036	10.030622	Home Economics	45	262	278.8667
9	100102 Visual Arts	100102	WA SENIOR HIGH SCHOOL WA	Wa Municipal	-2.28503036	10.030622	Visual Arts	45	250	275.2000
10	100104 General Arts	100104	LASSIE-TUOLO SNR SENIOR HIGH. SCHOOL. LASSIE	Wa Municipal	-2.28503036	10.030622	General Arts	45	319	337.4444

Exercise 3

1. The distance is created is in the Ldatstu_matching_sss_jss dataset, denoted as variable "distance".

2. Here provides distances of choices of student 340823 (the last one), in the order from his first choice to his sixth choice.

102.313410
189.327646
96.219306
96.219306
102.313410
23.507140

Exercise 4

1. The cutoff score of each "choice_rev" is provided in the "lowscore_rev" dataset. Here is the head of the dataset, with 10 first observations:

	choice_rev	minscore
1	100 arts	185
2	100 economics	188
3	100 others	185
4	100 science	198
5	101 arts	194
6	101 economics	193
7	101 others	196
8	101 science	202
9	102 arts	197
10	102 economics	197

2. The average score of each "choice_rev" is provided in the "quality_rev" dataset. Here is the head of the dataset, with 10 first observations:

	choice_rev	meanscore
1	1	265.5076
2	2	258.6949
3	3	253.4292
4	4	287.3222
5	5	310.5019
6	6	307.3822
7	7	288.7584
8	8	349.5031
9	9	309.1769
10	10	308.6568

Exercise 5

1. I keep data with students with their first choice. Since each student's score in invariant to their choices, in this question, I use the multinomial logit model.

2. Even after considering the 20000 students with highest scores, the dataset is still too large to get results. Therefore, I choose a smaller sample called "small_sample" with 300 observations with 60 unique choices and do the following assignment with this smaller sample. I re-factor these choices to estimate coefficients and marginal effects.

3. The coefficient of this model is shown in the following figure. The "c_constant" are constant coefficients for choices factored as 2 to 60 (the first choice normalized to 0) while "c_score" are coefficients for students' scores for choices factored as 2 to 60 (the first choice normalized to 0).

	c_constant	c_score
1	0.00000	0.00000000
2	-41.41242	0.11568715
3	-40.67462	0.11230960
4	45.34237	-0.12968341
5	-53.94772	0.14800848
6	-43.03350	0.11645097
7	79.86902	-0.22604521
8	-39.48290	0.10291150
9	253.29576	-0.71169969
10	-46.84558	0.12231285

4. The coefficient of this model is shown in the following figure. Here I only provide the head of it with the first 10 marginal effects.

	me
1	-0.000521401321
2	0.000464551070
3	0.000226257360
4	-0.000349239089
5	0.001210910525
6	0.000127296753
7	-0.000547979461
8	0.000010361926
9	-0.001929225848
10	0.000033612835

Exercise 6

1. For exercise 6, the quality of a school varies with each school, so I use the conditional logit model.

2. I still use the small sample with 300 observations and 60 unique choices.

3. The coefficient of the quality of school is approximately 0.014.

4. The constant coefficients is vector "co-constant" and are shown in the following figure (the first choice normalized to 0).

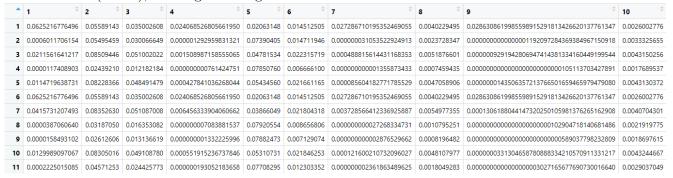|	co_constant
1	0.00000000
2	3.03484853
3	2.56121676
4	-0.30570327
5	2.26441541
6	1.02731195
7	-0.32559387
8	-1.16603298
9	-0.60820879
10	-0.60106241

5. The coefficient of this model is shown in the following figure. Here only provides the head of it (10*10), the original "clogit_me" matrix is 60*60.

	1	2	3	4	5	6	7	8	9	10
1	0.0000461701510	-0.000002933444	-0.000001698318	-0.0000001544298	-0.000002780106	-0.0000007722970	-0.0000001544289	-0.0000001544727	-0.0000001544429	-0.0000001544343
2	-0.0000029334436	0.000821220197	-0.000032141934	-0.0000029226983	-0.000052615558	-0.0000146162926	-0.0000029226812	-0.0000029235114	-0.0000029229463	-0.0000029227832
3	-0.0000016983183	-0.000032141934	0.000488979108	-0.0000016920973	-0.000030461799	-0.0000084621084	-0.0000016920875	-0.0000016925681	-0.0000016922409	-0.0000016921465
4	-0.0000001544298	-0.000002922698	-0.000001692097	0.0000460015936	-0.000002769922	-0.0000007694680	-0.0000001538632	-0.0000001539069	-0.0000001538772	-0.0000001538686
5	-0.0000027801055	-0.000052615558	-0.000030461799	-0.0000027699219	0.000781043398	-0.0000138522641	-0.0000027699057	-0.0000027706924	-0.0000027701569	-0.0000027700024
6	-0.0000007722970	-0.000014616293	-0.000008462108	-0.0000007694680	-0.000013852264	0.0002269734471	-0.0000007694635	-0.0000007696821	-0.0000007695333	-0.0000007694904
7	-0.0000001544289	-0.000002922681	-0.000001692087	-0.0000001538632	-0.000002769906	-0.0000007694635	0.0000460013263	-0.0000001539060	-0.0000001538763	-0.0000001538677
8	-0.0000001544727	-0.000002923511	-0.000001692568	-0.0000001539069	-0.000002770692	-0.0000007696821	-0.0000001539060	0.0000460143480	-0.0000001539200	-0.0000001539114
9	-0.0000001544429	-0.000002922946	-0.000001692241	-0.0000001538772	-0.000002770157	-0.0000007695333	-0.0000001538763	-0.0000001539200	0.0000460054837	-0.0000001538816
10	-0.0000001544343	-0.000002922783	-0.000001692147	-0.0000001538686	-0.000002770002	-0.0000007694904	-0.0000001538677	-0.0000001539114	-0.0000001538816	0.0000460029264

Exercise 7

1. The model in Exercise 5 is better than that in exercise 6 because each school's quality is computed by data of students and may suffers from endogeneity problem. Also, when making choices, students may view the quality of schools as fixed and consider based on his own scores.

2. The choice probability matrix for model in exercise 5 is the "prob_matrix". Here only provides the head of it (10*10), the original "clogit_me" matrix is 300*60.

	1	2	3	4	5	6	7	8	9	10
1	0.0625216776496	0.05589143	0.035002608	0.024068526805661950	0.02063148	0.014512505	0.027286710195352469055	0.0040229495	0.0286308619985598915291813426620137761347	0.0026002776
2	0.0006011706154	0.05495459	0.030066649	0.000001292959831321	0.07390405	0.014711946	0.000000031053522924913	0.0023728347	0.0000000000000001192097284369384967150918	0.0033325655
3	0.0211561641217	0.08509446	0.051002022	0.001508987158555065	0.04781534	0.022315719	0.000488815614431168353	0.0051876601	0.0000009291942806947414381334160449199544	0.0043150256
4	0.0000117408903	0.02439210	0.012182184	0.000000000761424751	0.07850760	0.006666100	0.000000000001355873433	0.0007459435	0.0000000000000000000000000105113703427891	0.0017689537
5	0.0114719638731	0.08228366	0.048491479	0.000427841036268044	0.05434560	0.021661165	0.000085604182771785529	0.0047058906	0.0000000143506357213766501659465979479080	0.0043130372
6	0.0625216776496	0.05589143	0.035002608	0.024068526805661950	0.02063148	0.014512505	0.027286710195352469055	0.0040229495	0.0286308619985598915291813426620137761347	0.0026002776
7	0.0415731207493	0.08352630	0.051087008	0.006456333904060662	0.03866049	0.021804318	0.003728566412336925887	0.0054977355	0.0001306188044147320250105981376265162908	0.0040704301
8	0.0000387060640	0.03187050	0.016353082	0.000000007083881537	0.07920554	0.008656806	0.000000000027268334731	0.0010795251	0.0000000000000000000000102904718140681486	0.0021919775
9	0.0000158493102	0.02612606	0.013136619	0.000000001332225996	0.07882473	0.007129074	0.000000000002876529662	0.0008196482	0.0000000000000000000000589037798232809	0.0018697615
10	0.0129989097067	0.08305016	0.049108780	0.000551915236737846	0.05310731	0.021846253	0.000121600210732096027	0.0048107977	0.0000000331304658780888334210570911331217	0.0043244667
11	0.0002225015085	0.04571253	0.024425773	0.000000193052183658	0.07708295	0.012303352	0.000000002361863489625	0.0018049283	0.0000000000000000000302716567769730016640	0.0029037049

3. After removing choice others, it would become. The output matrix for this question is "prob_matrix_noother".
Here only provides the head of it:

	1	2	3	4	5
1	0.095833743098	0.06626730	0.041112961	0.02485364	0.017671953
2	0.001306052101	0.05663909	0.031302710	0.07641450	0.014732390
3	0.032942297457	0.08743424	0.052244617	0.04971210	0.023128508
4	0.000036023771	0.02552543	0.013048013	0.08171682	0.006528612
5	0.018898047847	0.08414271	0.049556400	0.05614268	0.022188555