ECON613
Prof. Modibo Sidibe
Spring 2022

Homework 3: Data and multinomial choices

Runling Wu

1 Exercise 1 Basic Statistics

- There are 340823 students, 1609 schools and 33 unique programs. To calculate the number of programs, we pool choicepgm1 and choicepgm6 as a separate dataset and calculate the number of unique programs on top of schools.
- The total number of choices (school, program) matrix (account for repeated choices) is 2044938, if same school and same program only accounts once, we have 3086 unique school-program choices bundle.
- There are 265464 students applying to at least one senior high schools in the same district to home.
- Number of students each senior high school admitted shows in following table. I only include the first 15 rows, the rest is very similar. Code in r file.

	value	n
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
11	10111	600
12	10112	300
13	10114	350
14	10115	238
15	10116	446

- The cutoff of senior high schools (the lowest score to be admitted) shows in the following table. Only include first 15 rows, more cutoff info in the r code.
- The quality of senior high schools (the average score of students admitted).

	school_choice	cutoff	quality	size
1	10101	284	320.23	398
2	10102	343	394.15	248
3	10103	316	353.83	443
4	10104	245	296.92	220
5	10105	260	351.21	346
6	10106	293	340.10	395
7	10107	281	311.95	306
8	10108	248	303.90	318
9	10109	257	281.82	300
10	10110	343	408.08	535
11	10111	371	412.51	600
12	10112	316	375.61	300
13	10114	319	346.22	350
14	10115	274	316.34	238
15	10116	205	289.96	446

2 Exercise 2 Data

See the following first 10 rows of the school-level dataset, each row corresponding to (school, program).

	schoolcode	school_program	size	cutoff	quality	ssslong	sssdistrict
1	10101.00	10101 Agriculture	588	288	310.14	-0.20	Accra Metropolitan
2	10101.00	10101 Business	1200	305	324.86	-0.20	Accra Metropolitan
3	10101.00	10101 General Arts	1200	316	330.09	-0.20	Accra Metropolitan
4	10101.00	10101 General Science	600	299	329.10	-0.20	Accra Metropolitan
5	10101.00	10101 Home Economics	588	284	300.57	-0.20	Accra Metropolitan
6	10101.00	10101 Visual Arts	600	296	311.54	-0.20	Accra Metropolitan
7	10102.00	10102 General Arts	704	388	404.98	-0.20	Accra Metropolitan
8	10102.00	10102 General Science	560	389	406.41	-0.20	Accra Metropolitan
9	10102.00	10102 Home Economics	360	363	377.11	-0.20	Accra Metropolitan
10	10102.00	10102 Visual Arts	360	343	370.93	-0.20	Accra Metropolitan

3 Exercise 3 Distance

The following table shows the calculated distance using the formula. The variable school_program denotes the choice (school, program).

	V1.x	distance	school_program
1	1	8.81	50112 Home Economics
7	2	0.00	70102 General Arts
14	3	0.00	50702 Business
26	4	0.00	90501 Visual Arts
38	5	102.39	51802 Home Economics
50	6	121.57	10102 General Arts
58	7	27.48	80301 General Arts
67	8	34.22	40301 General Arts
77	9	0.00	21303 Business
89	10	138.74	80101 General Arts

4 Exercise 4 Dimensionality Reduction

For this question, I use two methods. From method 1, after reducing dimensions, I have the following dataset.

	V1	choice	choice_rev	score	size	cutoff	quality
1	1	1.00	501 Economics		2415	207	349.14
2	1	4.00	502 Arts		1420	205	317.15
3	1	6.00	509 Arts		188	216	256.20
4	1	3.00	502 Arts		1420	205	317.15
5	1	5.00	507 Economics		1083	212	275.27
6	1	2.00	501 Arts		2948	259	358.70
7	2	6.00	706 Arts		678	202	286.87
8	2	4.00	701 Arts		1161	202	297.11
9	2	1.00	701 Arts		1161	202	297.11
10	2	5.00	706 Economics		443	202	291.76

For the second method, I have this following dataset.

	choice	cutoff	quality	size
1	501Economics	207	349.14	2415
2	701Arts	202	297.11	1161
3	507Economics	212	275.27	1083
4	905Arts	194	286.67	265
5	518Economics	216	296.83	590
6	101Arts	243	340.08	3201
7	803Arts	218	262.75	350
8	403Arts	209	266.72	436
9	213Economics	201	304.78	705
10	801Arts	210	292.76	918

If we compare the first row, for schoolcode+programcode, 501Economics, two methods generate the same cutoff, quality, size.

5 Exercise 5 First Model

- Since students' test score are different across students, thus I will use a multinomial logit model to understand the effect of the student test score on his first choice.
- Use multinomial logit model in EX5, because test score is not the same for households. The last half coefficients here are the effect of test score(choice 2 to last choice), which mean that comparing to the probability of the first observed choice 1, individual will more likely to choose that choice if positive, and less likely to choose that choice if negative. The first half coefficients are the intercepts(choice 2 to last choice).
- see results in R.

6 Exercise 6 Second Model

- Since school quality is same across students, thus I will use a **conditional logit model** to understand the effect of the school quality on his first choice.
- The last coefficient here means that school quality and choice are negatively related– that is, higher school quality results in a less probability for the program to be chose. Each means that comparing to choice 1, an individual is more likely to choose that choice if the coefficient is positive, and less likely to choose that choice if the coefficient is negative.
- see results in R.

7 Exercise 7 Counterfactual simulations

- I think second model is appropriate to conduct this exercise. Both models are suitable for this setting. I choose conditional logit model to do the following exercise.
- see results in R.