Shuhan Zeng Econ 613-A1

Part 1

Exercise 1

1. Number of students
   340823
2. Number of schools
   640
3. Number of programs
   32
4. Number of choices
   2773
5. Number of students without test scores
   179887
6. Number of students applying to different programs within the same school
   120071
7. Number of students applying to less than 6 choices

Case 1: assume a student applies to less than 6 choices when the total number of schools and programs the student applies to is less than 12

21001

Case 2: assume a student applies to less than 6 choices when the number of applied schools is less than 6

17734

Case 3: assume a student applies to less than 6 choices when the number of applied programs is less than 6

20988

Exercise 2

	schoolcode	choicepgm	sssdistrict	ssslong	ssslat	cutoff	quality	size
1	10101	Agriculture	Accra Metropolitan	-0.1971153	5.607396	288	310.1429	49
2	10101	Business	Accra Metropolitan	-0.1971153	5.607396	305	324.8600	100
3	10101	General Arts	Accra Metropolitan	-0.1971153	5.607396	316	330.0900	100
4	10101	General Science	Accra Metropolitan	-0.1971153	5.607396	299	329.1000	50
5	10101	Home Economics	Accra Metropolitan	-0.1971153	5.607396	284	300.5714	49
6	10101	Technical	Accra Metropolitan	-0.1971153	5.607396	*NA*	*NA*	*NA*
7	10101	Visual Arts	Accra Metropolitan	-0.1971153	5.607396	296	311.5400	50
8	10102	General Arts	Accra Metropolitan	-0.1971153	5.607396	388	404.9773	88
9	10102	General Science	Accra Metropolitan	-0.1971153	5.607396	389	406.4143	70
10	10102	Home Economics	Accra Metropolitan	-0.1971153	5.607396	363	377.1111	45
11	10102	Visual Arts	Accra Metropolitan	-0.1971153	5.607396	343	370.9333	45
12	10103	Agriculture	Accra Metropolitan	-0.1971153	5.607396	316	333.1316	38
13	10103	Business	Accra Metropolitan	-0.1971153	5.607396	341	357.9664	119

	schoolcode	choicepgm	sssdistrict	ssslong	ssslat	cutoff	quality	size
14	10103	General Arts	Accra Metropolitan	-0.1971153	5.607396	349	362.5812	117
15	10103	General Science	Accra Metropolitan	-0.1971153	5.607396	335	353.5625	80
16	10103	Home Economics	Accra Metropolitan	-0.1971153	5.607396	320	336.0408	49
17	10103	Visual Arts	Accra Metropolitan	-0.1971153	5.607396	343	357.9500	40
18	10104	General Arts	Accra Metropolitan	-0.1971153	5.607396	302	320.1273	55
19	10104	General Science	Accra Metropolitan	-0.1971153	5.607396	245	283.3636	55
20	10104	Home Economics	Accra Metropolitan	-0.1971153	5.607396	264	285.8545	55

Exercise 3

The whole dataset is too large to be reported, so I only report the first 20 rows and 4 columns here.

	schoolcode	sssdistrict	jssdistrict	distance
1	30107	Cape Coast Municipal	South Dayi (Kpeve)	439.1766
2	30107	Cape Coast Municipal	South Dayi (Kpeve)	439.1766
3	30103	Cape Coast Municipal	South Dayi (Kpeve)	439.1766
4	30107	Cape Coast Municipal	South Dayi (Kpeve)	439.1766
5	30103	Cape Coast Municipal	South Dayi (Kpeve)	439.1766
6	30103	Cape Coast Municipal	South Dayi (Kpeve)	439.1766
7	21003	Kwahu South (Mpraeso)	South Dayi (Kpeve)	430.5811
8	10111	Ga West (Amasaman)	South Dayi (Kpeve)	428.6852
9	30104	Cape Coast Municipal	South Dayi (Kpeve)	439.1766
10	30107	Cape Coast Municipal	South Dayi (Kpeve)	439.1766
11	20301	Akwapim South (Nsawam)	South Dayi (Kpeve)	427.9143
12	21003	Kwahu South (Mpraeso)	South Dayi (Kpeve)	430.5811
13	50110	Kumasi Metro	South Dayi (Kpeve)	444.3355
14	10110	Accra Metropolitan	South Dayi (Kpeve)	427.5683
15	30103	Cape Coast Municipal	South Dayi (Kpeve)	439.1766
16	40103	Shama/Ahanta/East (Sekondi/Takoradi)	South Dayi (Kpeve)	444.8470
17	21103	Kwaebibirem (Kade)	South Dayi (Kpeve)	432.2377
18	20301	Akwapim South (Nsawam)	South Dayi (Kpeve)	427.9143
19	50102	Kumasi Metro	South Dayi (Kpeve)	444.3355
20	30107	Cape Coast Municipal	South Dayi (Kpeve)	439.1766

Exercise 4

1. The average and sd for each ranked choice
   First choice:

	cutoff	quality	distance
Stand dev	54.12556	48.5386	126.2553
Mean	294.23718	316.6898	533.3504

Second choice:

	cutoff	quality	distance
Stand dev	49.57892	43.82404	126.2991
Mean	281.35659	304.58320	533.3108

Third choice:

	cutoff	quality	distance
Stand dev	46.96337	41.16625	126.2993
Mean	272.68599	296.78969	533.2431

Fourth choice:

	cutoff	quality	distance
Stand dev	45.1399	39.2662	126.2757
Mean	263.0766	288.5767	533.1415

Fifth choice:

	cutoff	quality	distance
Stand dev	32.06955	26.73123	127.5923
Mean	250.08489	277.50407	535.4284

Sixth choice:

	cutoff	quality	distance
Stand dev	31.44553	26.23097	127.5850
Mean	246.18031	274.07874	535.5617

2. The average and sd differentiated by student test score quantiles

Group	Frist Quantile	Second Quantile	Third Quantile	Fourth Quantile	NA
Test score	[158,252]	(252,283]	(283,324]	(324,469]	NA's

Quantcut: Group by student score quantiles
NA: the group in which the values of student test score are missing

First choice:

	quantcut	cutoff_mean	quality_mean	dist_mean	cutoff_sd	quality_sd	dist_sd
1	[158,252]	283.4534	306.7529	547.8415	44.59166	38.94246	138.35497
2	(252,283]	301.0173	322.7784	535.6282	45.03198	39.25254	123.63237
3	(283,324]	323.0130	342.9473	521.5170	43.53460	38.32348	113.13859
4	(324,469]	362.1617	380.4723	494.2369	38.09107	34.85716	99.77107
5	NA	273.7193	297.6127	540.7815	47.34754	41.62563	130.00309

Second choice:

	quantcut	cutoff_mean	quality_mean	dist_mean	cutoff_sd	quality_sd	dist_sd
1	[158,252]	270.0009	294.3333	547.8249	41.40760	35.90342	138.36331
2	(252,283]	285.1532	308.0621	535.5875	42.37029	36.63525	123.63509
3	(283,324]	304.1427	325.3159	521.4517	42.36073	36.63066	113.17221
4	(324,469]	340.5253	358.9266	493.9973	38.24956	33.94823	99.98297
5	NA	264.8867	289.4607	540.7868	44.16029	38.58289	130.02216

Third choice:

	quantcut	cutoff_mean	quality_mean	dist_mean	cutoff_sd	quality_sd	dist_sd
1	[158,252]	261.2568	286.6283	547.7784	40.43917	34.93599	138.3506
2	(252,283]	273.9051	298.0751	535.5236	41.35209	35.49417	123.6023
3	(283,324]	290.5017	313.0342	521.3369	41.65390	35.69395	113.1395
4	(324,469]	324.1210	343.6262	493.7254	39.78167	34.51467	100.0950
5	NA	259.6227	284.7935	540.7702	42.80620	37.32703	130.0024

Fourth choice:

	quantcut	cutoff_mean	quality_mean	dist_mean	cutoff_sd	quality_sd	dist_sd
1	[158,252]	251.4020	278.2927	547.6526	39.75034	34.20236	138.3196
2	(252,283]	262.0434	288.0557	535.4071	40.85982	34.66926	123.5365
3	(283,324]	276.2045	300.7063	521.2626	41.29130	34.98959	113.0902
4	(324,469]	306.9100	328.3540	493.4150	41.69904	35.64904	100.0990
5	NA	253.3101	279.5001	540.7107	42.12317	36.66327	129.9705

Fifth choice:

	quantcut	cutoff_mean	quality_mean	dist_mean	cutoff_sd	quality_sd	dist_sd
1	[158,252]	246.8855	274.3982	549.7461	31.22731	25.74618	139.4084
2	(252,283]	253.4241	280.8732	537.6023	31.44824	25.36836	124.8482
3	(283,324]	260.0547	287.4211	523.2893	31.18633	24.47648	114.4673
4	(324,469]	266.4038	294.6718	496.5178	30.28669	22.71537	101.0742
5	NA	244.3747	271.5945	542.6888	31.13064	26.18481	131.3455

Sixth choice:

	quantcut	cutoff_mean	quality_mean	dist_mean	cutoff_sd	quality_sd	dist_sd
1	[158,252]	242.4731	270.5157	549.8932	30.73168	25.52473	139.3961
2	(252,283]	248.7229	276.7619	537.6978	31.12694	25.21741	124.8185
3	(283,324]	254.6174	282.5944	523.3684	31.21378	24.75947	114.4261
4	(324,469]	260.5411	289.3772	496.8183	30.53754	23.10501	101.0953
5	NA	241.5311	269.1465	542.7761	30.53912	25.74210	131.3457

Part 2

Exercise 5

The first 20 observations:

	X1	X2	X3	error	Y	ydum	X0
1	1.883715	6.1626934	0	1.3511829	-1.4347830	0	1
2	1.947702	0.8184512	1	2.0805542	4.2811902	1	1
3	1.665419	10.0422301	0	1.9736556	-4.5658483	0	1
4	2.201522	7.7359735	0	2.4894315	-1.3311179	0	1
5	1.815048	14.1547923	0	0.5377489	-9.5235068	0	1
6	1.664311	7.4204797	1	1.7412995	-2.3399585	0	1
7	2.506237	3.3859738	0	2.9110203	3.3711281	1	1
8	2.564542	2.2236426	0	3.1024779	4.6786498	1	1
9	2.057414	5.0575344	1	-0.6559440	-2.1388283	0	1
10	1.111271	3.2160337	0	3.5601251	2.4992199	1	1
11	1.244696	2.8513225	0	3.4119477	2.8393926	1	1

	X1	X2	X3	error	Y	ydum	X0
12	1.256017	17.9571471	1	0.4212251	-13.6329865	0	1
13	2.306041	5.1824378	0	3.0048351	1.6078908	1	1
14	2.445641	7.7754909	1	1.7197170	-1.7434552	0	1
15	1.289525	10.6069578	0	3.1527616	-4.3460708	0	1
16	1.451560	1.0926519	0	2.1963568	3.4548416	1	1
17	2.375549	11.8721781	1	2.6558598	-4.5784422	0	1
18	2.981613	6.6776150	0	1.0500587	-0.8818595	0	1
19	2.947235	3.5725781	0	2.0538133	2.8751751	1	1
20	1.288224	8.6597947	0	2.0427681	-3.7051786	0	1

Exercise 6

1. Correlation between Y and X1:
   The correlation between Y and X1 is 0.1994163 (always around 0.2) and it is different from 1.2.
   $$\hat{\beta} = cor(Y_i, X_i) * \frac{sd(Y_i)}{sd(X_i)}$$
   The correction between Y and X1 measures the strength of the linear relationship between Y and X1, while beta shows the expected change in Y given a unit change in X1. Since $sd(Y_i)$ is not equal to $sd(X_i)$, $cor(Y_i, X_i)$ is different from 1.2

2. Outcome of the regression of Y on X

	betas	SE
X0	2.48452742	0.04033776
X1	1.20670445	0.01723276
X2	-0.89886599	0.00285418
X3	0.09228005	0.02167972

Exercise 7

Glm:est: the estimation got from GLM function
Glm:se: se got from GLM function
own:est: the estimation got from likelihood maximization
own:se: se got from likelihood maximization

1. Probit

	glm : est	glm :se	own : est	own :se
(Intercept)	3.05595545	0.10078649	3.05597823	0.10061489

	glm : est	glm :se	own : est	own :se
X1	1.18139936	0.04381088	1.18139914	0.04382298
X2	-0.90854411	0.01860212	-0.90854823	0.01854023
X3	0.09061841	0.04711902	0.09062326	0.04714863

2. Logit

	glm : est	glm :se	own : est	own :se
(Intercept)	5.5045815	0.18980882	5.5045960	0.18981561
X1	2.1206590	0.08151260	2.1206656	0.08151532
X2	-1.6361753	0.03714629	-1.6361800	0.03714833
X3	0.1760413	0.08499998	0.1760422	0.08500180

3. Linear Probability

	lm : est	lm :se	own : est	own :se
(Intercept)	0.9097251	0.0133721963	0.9097251	0.0133721963
X1	0.1362626	0.0057127581	0.1362626	0.0057127581
X2	-0.1043247	0.0009461768	-0.1043247	0.0009461768
X3	0.0156355	0.0071869507	0.0156355	0.0071869507

4. Comparison

   According to the above results, the estimated coefficients for each variable in the three models have completely different values but the same sign. Probit model and logit model yield the same result but the different estimated parameters. The estimated coefficients of logit and probit model tell us that an increase in X1 and X3 would cause an increase in the probability that Y is higher than its mean, while an increase in X2 would cause a decrease in the probability. The results of linear probability model show that the predicted ydum would increase by around 0.1363 given a unit increase in X1, decrease by 0.1043 given a unit increase in X2, and increase by around 0.0156 given a unit increase in X3 (when X3=1). Since the predicted ydum in linear Probability model would not be between 0 and 1, the result of linear probability model might be inefficient.

   According to the computed standard errors for each coefficient, I believe the estimated coefficients for X1, X2, X3, X0 are all significant under 0.05 significance level.

Exercise 8

1. Average marginal effect

	X0	X1	X2	X3
probit	0.3616792	0.1398202	-0.1075279	0.01072539
logit	0.3604720	0.1388813	-0.1071493	0.01153009

2. Standard error by bootstrap
   Probit:

	X0	X1	X2	X3
Stand dev	0.09817298	0.09160822	0.03511453	0.02548894
Mean	0.32648086	0.15687578	-0.10848959	0.01710063

   Logit:

	X0	X1	X2	X3
Stand dev	0.008666987	0.004051598	0.0003828049	0.00575684
Mean	0.362079836	0.138135057	-0.1070946531	0.01178140