A3

Rulin Wang

Exercise1

(1) #340823 students
##640 schools
32 programs

(2) ## 2773 choices.

	schoolcode	choicepgm1
1	10101	Home Economics
2	10101	General Arts
3	10101	Business
4	10101	Visual Arts
5	10101	Agriculture
6	10101	General Science
7	10101	Technical
8	10102	General Arts
9	10102	Home Economics
10	10102	General Science
11	10102	Visual Arts
12	10103	General Arts

(3) ##265464 students

(4) "table"

	school	Freq
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

(5) “table1”

schoolcode	cutoff
10101	284
10102	343
10103	316
10104	245
10105	260
10106	293
10107	281
10108	248
10109	257
10110	343
10111	371
10112	316
10114	319

(6) “table2”

	schoolcode	quality
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

Exercise2

"data5"

	schoolcode	choicepgm1	sssdistrict	ssslong	ssslat	cutoff	quality	size
1	10101	Home Economics	Accra Metropolitan	-0.1971153	5.607396	284	300.5714	49
2	10101	General Arts	Accra Metropolitan	-0.1971153	5.607396	316	330.0900	100
3	10101	Business	Accra Metropolitan	-0.1971153	5.607396	305	324.8600	100
4	10101	Visual Arts	Accra Metropolitan	-0.1971153	5.607396	296	311.5400	50
5	10101	Agriculture	Accra Metropolitan	-0.1971153	5.607396	288	310.1429	49
6	10101	General Science	Accra Metropolitan	-0.1971153	5.607396	299	329.1000	50
7	10101	Technical	Accra Metropolitan	-0.1971153	5.607396	NA	NA	NA
8	10102	General Arts	Accra Metropolitan	-0.1971153	5.607396	388	404.9773	88
9	10102	Home Economics	Accra Metropolitan	-0.1971153	5.607396	363	377.1111	45
10	10102	General Science	Accra Metropolitan	-0.1971153	5.607396	389	406.4143	70
11	10102	Visual Arts	Accra Metropolitan	-0.1971153	5.607396	343	370.9333	45
12	10103	General Arts	Accra Metropolitan	-0.1971153	5.607396	349	362.5812	117
13	10103	Business	Accra Metropolitan	-0.1971153	5.607396	341	357.9664	119

Exercise3

dis1-dis6 in data8

lat6	dist1	dist2	dist3	dist4	dist5	dist6
462874	8.813579	8.813579	18.895053	18.895053	17.179653	63.91775
896852	0.000000	21.672792	0.000000	0.000000	21.672792	21.67279
806778	0.000000	0.000000	9.439135	0.000000	12.519350	0.00000
557073	0.000000	25.651061	70.461574	25.651061	70.461574	25.70067
681337	102.388006	42.229396	26.910431	26.910431	42.229396	26.91043
707927	121.565099	0.000000	8.813421	14.535402	22.380815	14.43245
916286	27.483623	0.000000	27.483623	0.000000	115.949734	78.91660
141226	34.220915	0.000000	0.000000	34.220915	83.022869	34.22092
112613	0.000000	0.000000	19.051111	19.051111	16.251337	11.71034
199565	138.742844	229.308384	22.311404	0.000000	22.311404	22.31140
462874	23.060819	28.747633	0.000000	28.747633	28.747633	22.31581
924120	110.447129	0.000000	109.583518	0.000000	50.558602	29.20631
617353	17.700575	17.700575	0.000000	0.000000	17.700575	0.00000

Exercise4

(1) scode_rev1-6 in data9

scode_rev1	scode_rev2	scode_rev3	scode_rev4	scode_rev5	scode_rev6
501	501	502	502	507	509
701	706	701	701	706	706
507	507	501	507	516	507
905	904	901	909	901	903
518	517	502	502	516	502
101	501	517	502	506	516
803	804	803	804	805	809
403	404	404	403	402	403
213	213	212	212	202	201
801	904	505	509	505	505
518	506	505	506	506	509
100	905	801	905	902	906
306	306	309	309	306	309

(2) pgm_rev1-6 in data9

pgm_rev1	pgm_rev2	pgm_rev3	pgm_rev4	pgm_rev5	pgm_rev6
economics	arts	arts	arts	economics	arts
arts	economics	arts	arts	economics	arts
economics	economics	economics	economics	economics	economics
arts	arts	others	others	others	arts
economics	arts	economics	arts	arts	economics
arts	arts	arts	arts	economics	economics
arts	arts	arts	arts	arts	arts
arts	arts	arts	others	others	others
economics	economics	science	science	arts	arts
arts	arts	arts	arts	arts	arts
economics	arts	economics	economics	arts	arts
science	science	science	science	science	others

(3) choice_rev1-6 in data9

choice_rev1	choice_rev2	choice_rev3	choice_rev4	choice_rev5	choice_rev6
501 economics	501 arts	502 arts	502 arts	507 economics	509 arts
701 arts	706 economics	701 arts	701 arts	706 economics	706 arts
507 economics	507 economics	501 economics	507 economics	516 economics	507 economics
905 arts	904 arts	901 others	909 others	901 others	903 arts
518 economics	517 arts	502 economics	502 arts	516 arts	502 economics
101 arts	501 arts	517 arts	502 arts	506 economics	516 economics
803 arts	804 arts	803 arts	804 arts	805 arts	809 arts
403 arts	404 arts	404 arts	403 others	402 others	403 others
213 economics	213 economics	212 science	212 science	202 arts	201 arts
801 arts	904 arts	505 arts	509 arts	505 arts	505 arts
518 economics	506 arts	505 economics	506 economics	506 arts	509 arts
100 science	905 science	801 science	905 science	902 science	906 others

(4) “table31” cutoff ; “table32” quality

	choice_rev	cutoff
1	100 arts	194
2	100 economics	195
3	100 others	191
4	100 science	228
5	101 arts	243
6	101 economics	205
7	101 others	257
8	101 science	203
9	102 arts	216
10	102 economics	206
11	102 others	209
12	102 science	242
13	103 arts	260

	choice_rev	quality
1	100 arts	275.5233
2	100 economics	264.4993
3	100 others	245.6381
4	100 science	305.1814
5	101 arts	340.0850
6	101 economics	326.3979
7	101 others	313.2753
8	101 science	368.7612
9	102 arts	315.5544
10	102 economics	308.9986
11	102 others	280.9509
12	102 science	340.3426
13	103 arts	299.1236

Exercise5

（1） Then we can build a model with dependent Variable: choice_rev (246 choices) and independent variables: test scores.

Because test scores are different for each student (choosers' character), we can use multinomial logit to analyze the effect of the student test score on his first choice.

```
likelyfun1 = function(beta, data, choice_number,
                      multinomial_v_num, multinomial_v_start) {
  N = nrow(data14)
  pij = mat.or.vec(N,choice_number)
  ch = data14$choice_rev
   pij[,1] = 1
    for(j in seq(1,choice_number-1))   ### omit a choice
      {pij[,j+1] = exp(beta[j] +
                          apply(data[,seq(multinomial_v_start,
                                  multinomial_v_start + multinomial_v_num - 1),
                             with=FALSE]*
                            beta[choice_number + j - 1 +
                                  choice_number   *   seq(0,multinomial_v_num-
1)],1,sum)) }
  prob    = sweep(pij,MARGIN=1,FUN="/",STATS=rowSums(pij))
  probc = NULL
  for (i in 1:N)
  {probc[i] = prob[i,ch[i]]}
  probc[probc>0.999999] = 0.999999
  probc[probc<0.000001] = 0.000001
  like = sum(log(probc))
  return(-like) }     ##likelyhood function
```

(2) “multi12”

	(Intercept)	score
2	1.159890e-01	1.129448e-03
3	-5.900832e-03	-1.500258e-03
4	2.314422e-01	3.570387e-03
5	1.201838e+00	6.753547e-03
6	1.335635e+00	2.681114e-03
7	-9.099455e-05	2.726235e-03
8	-9.254632e+00	3.454959e-02
9	5.382385e-01	2.654350e-03
10	4.378082e-01	3.451623e-03
11	1.958375e-03	3.775479e-04
12	1.976038e-01	2.703180e-03
13	2.208058e-02	-3.313528e-03
14	6.099487e-03	-2.066087e-03

“Me1”

	marginal effect
X1	-2.593355e-06
X2	-4.479558e-06
X3	-1.461515e-06
X4	-1.273326e-05
X5	-1.134417e-04
X6	-2.740289e-05
X7	-7.311530e-06
X8	-1.597432e-04
X9	-1.220696e-05
X10	-1.493491e-05
X11	-3.008195e-06
X12	-8.837831e-06
X13	-7.553183e-07

Exercise6

(1) # Then we need to bulid a model with dependent Variable: choice_rev (246 choices) and independent variables: quality. Because quality is the character for each choice (same choices have same characters), we can use conditional logit to analyze the effect of the student test score on his first choice.

```
likelyfun2 = function(beta, data, choice_number,
                                    conditional_v_num, conditional_v_start) {
            N = nrow(data14)
            pij = mat.or.vec(N,choice_number)
            ch = data14$choice_rev
            pij[,1] = exp(0 + apply(data[,conditional_v_start +
                                                choice_number * seq(0,conditional_v_num-1),
                                              with=FALSE]*beta[seq(choice_number,

    choice_number+conditional_v_num-1)],
                                         1,sum))
            for(j in seq(1,choice_number-1)){
              pij[,j+1] = exp(beta[j] +
                                    apply(data[,conditional_v_start +
                                                  choice_number * seq(0,conditional_v_num-1)
    + j,
                                                with=FALSE] * beta[seq(choice_number,

    choice_number+conditional_v_num-1)],
                                           1,sum)) }
          prob = sweep(pij,MARGIN=1,FUN="/",STATS=rowSums(pij))
          probc = NULL
          for (i in 1:N)
          {probc[i] = prob[i,ch[i]]}
          probc[probc>0.999999] = 0.999999
          probc[probc<0.000001] = 0.000001
          like = sum(log(probc))
    return(-like)}
```

(2)

Exercise7

(1) I think the second model (conditional logit) is better, because “others” is a school characteristics rather than individual characteristics. Thus the effect to quality is less than scores when removing “choices”. (quality only reduces 1 character)

(2)

(2) Compare reg1 and reg2 can get the result, but I didn't figure it out. I guess choice probabilities do not change much, because there aren't many "others".