# Applied Microeconometrics - Assignment 4

483670, 630516

September 23, 2022

## 1

Table 1 contains the results of a multinomial logit model of school choice on gender and test scores.

**Table 1**

| choice | Coefficient | Std. err. | z | P>\|z\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| **High_school_1** | (base alternative) | | | | | |
| **Outside_Utrecht** | | | | | | |
| gender | −.4690375 | .4415235 | −1.06 | 0.288 | −1.334408 | .3963327 |
| testscore | −.0506644 | .0545106 | −0.93 | 0.353 | −.1575032 | .0561744 |
| _cons | 29.04348 | 29.6866 | 0.98 | 0.328 | −29.1412 | 87.22815 |
| **Gymnasium_1** | | | | | | |
| gender | −.419388 | .4363943 | −0.96 | 0.337 | −1.274705 | .435929 |
| testscore | .1200623 | .0566682 | 2.12 | 0.034 | .0089948 | .2311299 |
| _cons | −63.95478 | 30.89643 | −2.07 | 0.038 | −124.5107 | −3.398893 |
| **High_school_2** | | | | | | |
| gender | −.2547803 | 1.468061 | −0.17 | 0.862 | −3.132128 | 2.622567 |
| testscore | .2180528 | .2512237 | 0.87 | 0.385 | −.2743366 | .7104423 |
| _cons | −121.5995 | 137.5185 | −0.88 | 0.377 | −391.1309 | 147.9319 |
| **High_school_3** | | | | | | |
| gender | −.2168744 | .4658658 | −0.47 | 0.642 | −1.129955 | .6962058 |
| testscore | −.0206016 | .0578131 | −0.36 | 0.722 | −.1339133 | .0927101 |
| _cons | 12.13533 | 31.48937 | 0.39 | 0.700 | −49.5827 | 73.85335 |
| **High_school_4** | | | | | | |
| gender | −.0134729 | .5383483 | −0.03 | 0.980 | −1.068616 | 1.04167 |
| testscore | −.0068987 | .0669302 | −0.10 | 0.918 | −.1380795 | .1242822 |
| _cons | 3.837665 | 36.45847 | 0.11 | 0.916 | −67.61962 | 75.29495 |
| **High_school_5** | | | | | | |
| gender | −.1953817 | 1.468223 | −0.13 | 0.894 | −3.073045 | 2.682282 |
| testscore | −.0806672 | .1717818 | −0.47 | 0.639 | −.4173533 | .256019 |
| _cons | 41.40126 | 93.35539 | 0.44 | 0.657 | −141.5719 | 224.3745 |
| **High_school_6** | | | | | | |
| gender | −.2324838 | .4710565 | −0.49 | 0.622 | −1.155738 | .6907699 |
| testscore | .0354673 | .0596902 | 0.59 | 0.552 | −.0815233 | .1524579 |
| _cons | −18.48359 | 32.52928 | −0.57 | 0.570 | −82.23981 | 45.27262 |
| **High_school_7** | | | | | | |
| gender | .4913499 | .6720051 | 0.73 | 0.465 | −.8257559 | 1.808456 |
| testscore | −.0639538 | .0781985 | −0.82 | 0.413 | −.21722 | .0893124 |
| _cons | 33.92765 | 42.54658 | 0.80 | 0.425 | −49.46212 | 117.3174 |
| **Gymnasium_2** | | | | | | |
| gender | −.3593733 | .5188671 | −0.69 | 0.489 | −1.376334 | .6575876 |
| testscore | .0798507 | .0679752 | 1.17 | 0.240 | −.0533782 | .2130796 |
| _cons | −43.14291 | 37.07115 | −1.16 | 0.245 | −115.801 | 29.51521 |

## 2

Table 2 contains the results of a multinomial logit model of school choice on the distance to a given school and whether the deciding student has a sibling at this school. The results suggest that if your sibling is at the school your probability of choosing it increases (positive and significant coefficient). The negative significant coefficient for the distance suggests that the greater the distance to the school, the lower the probability of choosing that school. If we convert them to odds ratio, having a sibling at a school makes you just over 5 times more likely to choose that school - all else equal. For each kilometre a school is further away than another, it is 20% (not percentage points-careful) less likely to be chosen.

**Table 2**

| choice | Coefficient | Std. err. | z | P>\|z\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| **school** | | | | | | |
| sibling | 1.689151 | .1895964 | 8.91 | 0.000 | 1.317549 | 2.060753 |
| distance | -.245953 | .0298558 | -8.24 | 0.000 | -.3044694 | -.1874367 |
| **High_school_1** | (base alternative) | | | | | |
| **Outside_Utrecht** | | | | | | |
| _cons | 1.42745 | .2385271 | 5.98 | 0.000 | .9599457 | 1.894955 |
| **Gymnasium_1** | | | | | | |
| _cons | 1.236214 | .2324322 | 5.32 | 0.000 | .7806554 | 1.691773 |
| **High_school_2** | | | | | | |
| _cons | -2.142987 | .743227 | -2.88 | 0.004 | -3.599686 | -.6862892 |
| **High_school_3** | | | | | | |
| _cons | .9008862 | .2558284 | 3.52 | 0.000 | .3994717 | 1.402301 |
| **High_school_4** | | | | | | |
| _cons | -.152363 | .2778213 | -0.55 | 0.583 | -.6968827 | .3921566 |
| **High_school_5** | | | | | | |
| _cons | -2.614989 | .7388643 | -3.54 | 0.000 | -4.063136 | -1.166841 |
| **High_school_6** | | | | | | |
| _cons | .8443403 | .2568296 | 3.29 | 0.001 | .3409635 | 1.347717 |
| **High_school_7** | | | | | | |
| _cons | -.590173 | .3341246 | -1.77 | 0.077 | -1.245045 | .0646992 |
| **Gymnasium_2** | | | | | | |
| _cons | .2806044 | .2787649 | 1.01 | 0.314 | -.2657647 | .8269735 |

## 3

Table 3 combines the regressions of table 1 and table 2. This results in a rank-ordered logit model. The siblings' school and the distance to a given school

remain similarly important in determining future students' preferences.

**Table 3**

| choice | Coefficient | Std. err. | z | P>\|z\| | [95% conf. | interval] |
|---|---|---|---|---|---|---|
| **school** | | | | | | |
| sibling | 1.674634 | .1934232 | 8.66 | 0.000 | 1.295531 | 2.053737 |
| distance | -.2605671 | .0306829 | -8.49 | 0.000 | -.3207045 | -.2004297 |
| **High_school_1** | (base alternative) | | | | | |
| **Outside_Utrecht** | | | | | | |
| gender | -.6169134 | .4638332 | -1.33 | 0.184 | -1.52601 | .292183 |
| testscore | -.0255972 | .0567346 | -0.45 | 0.652 | -.136795 | .0856006 |
| _cons | 15.67982 | 30.88092 | 0.51 | 0.612 | -44.84567 | 76.20531 |
| **Gymnasium_1** | | | | | | |
| gender | -.5939652 | .4680242 | -1.27 | 0.204 | -1.511276 | .3233455 |
| testscore | .1496682 | .0600876 | 2.49 | 0.013 | .0318987 | .2674378 |
| _cons | -80.1084 | 32.74004 | -2.45 | 0.014 | -144.2777 | -15.93909 |
| **High_school_2** | | | | | | |
| gender | -.4408758 | 1.487665 | -0.30 | 0.767 | -3.356645 | 2.474894 |
| testscore | .2503491 | .2510607 | 1.00 | 0.319 | -.2417208 | .742419 |
| _cons | -138.6169 | 137.4542 | -1.01 | 0.313 | -408.0222 | 130.7884 |
| **High_school_3** | | | | | | |
| gender | -.6193152 | .5079289 | -1.22 | 0.223 | -1.614837 | .3762071 |
| testscore | .0266032 | .0628187 | 0.42 | 0.672 | -.0965193 | .1497257 |
| _cons | -13.27733 | 34.20047 | -0.39 | 0.698 | -80.30901 | 53.75436 |
| **High_school_4** | | | | | | |
| gender | -.0016342 | .5597209 | -0.00 | 0.998 | -1.098667 | 1.095399 |
| testscore | .0127403 | .0700957 | 0.18 | 0.856 | -.1246446 | .1501253 |
| _cons | -7.089617 | 38.17082 | -0.19 | 0.853 | -81.90304 | 67.72381 |
| **High_school_5** | | | | | | |
| gender | -.3748867 | 1.480941 | -0.25 | 0.800 | -3.277477 | 2.527704 |
| testscore | -.0264712 | .1755886 | -0.15 | 0.880 | -.3706186 | .3176761 |
| _cons | 11.94192 | 95.41748 | 0.13 | 0.900 | -175.0729 | 198.9567 |
| **High_school_6** | | | | | | |
| gender | -.4506662 | .5073866 | -0.89 | 0.374 | -1.445126 | .5437932 |
| testscore | .0923924 | .0641304 | 1.44 | 0.150 | -.0333009 | .2180857 |
| _cons | -49.27242 | 34.93046 | -1.41 | 0.158 | -117.7349 | 19.19003 |
| **High_school_7** | | | | | | |
| gender | .4013687 | .691612 | 0.58 | 0.562 | -.9541659 | 1.756903 |
| testscore | -.031933 | .0797927 | -0.40 | 0.689 | -.1883238 | .1244578 |
| _cons | 16.53 | 43.40518 | 0.38 | 0.703 | -68.5426 | 101.6026 |
| **Gymnasium_2** | | | | | | |
| gender | -.6693524 | .5540754 | -1.21 | 0.227 | -1.75532 | .4166153 |
| testscore | .1390897 | .0726138 | 1.92 | 0.055 | -.0032307 | .2814101 |
| _cons | -75.21529 | 39.58453 | -1.90 | 0.057 | -152.7995 | 2.368956 |

## 4

The theory of independence of irrelevant alternatives means that the coefficients and effects that we have predicted would remain the same if we remove one of the options. This is likely to be the case when all options are substitutes. Schools in our case are all substitutes for education. They may differ, but primarily on observables (type of school, rating, distance from decision maker). However, the regression we have run has not included all those observed. In the analysis that we have run not all are very good alternatives so not all other options are irrelevant. For example, gymnasiums are more similar to other gymnasium than no-gymnasium schools. In this situation removing a gymnasium is likely to shift those that were there disproportionately to the other gymnasiums, as opposed

to the predicted probabilities of our model. To help here it would be beneficial to include gymnasium in our regression as a variable. The same argument however can also be made for other variables that are either not measured here or are not measurable at all.

## Code - Stata

```
use "/Users/julian/Documents/Current/Applied Microeconometrics/GitHub/Assignment 4/HighSchooData.dta"
#Table 1
asclogit choice, case(id) alternatives(school) casevars(gender testscore)
#Table 2
asclogit choice sibling distance, case(id) alternatives(school)
#Table 3
asclogit choice sibling distance, case(id) alternatives(school) casevars(gender testscore)
```