Assignment 1: Data Manipulation - OLS/Discrete Choice

Ziyuan Wang

Part 1

Exercise 1 Missing data

```
##                                 [,1]
## number.of.students            340823
## number.of.schools                842
## number.of.programs                33
## number.of.choices               3086
## number.of.missing.test.score 179887
## apply.to.the.same.school     608970
## apply.to.less.than.6.choices  18954
```

Exercise 2 Data

```
##                  SchPrgm  sssdistrict   ssslong   ssslat Cutoff  Quality Size
## 1     100101General Arts Wa Municipal -2.285030 10.03062    198 244.3924   79
## 2   100101Home Economics Wa Municipal -2.285030 10.03062    199 229.4500   40
## 3        100101Technical Wa Municipal -2.285030 10.03062    201 235.1020   49
## 4      100102Agriculture Wa Municipal -2.285030 10.03062    273 292.5556   90
## 5         100102Business Wa Municipal -2.285030 10.03062    283 303.3444   90
## 6     100102General Arts Wa Municipal -2.285030 10.03062    291 311.1111   90
## 7  100102General Science Wa Municipal -2.285030 10.03062    273 298.4333   90
## 8   100102Home Economics Wa Municipal -2.285030 10.03062    262 278.8667   45
## 9      100102Visual Arts Wa Municipal -2.285030 10.03062    250 275.2000   45
## 10    100104General Arts Wa Municipal -2.285030 10.03062    319 337.4444   45
## 11 100104General Science Wa Municipal -2.285030 10.03062    313 334.0000   45
## 12  100104Home Economics Wa Municipal -2.285030 10.03062    282 309.3556   45
## 13        100105Business Wa Municipal -2.285030 10.03062    251 268.0125   80
## 14    100105General Arts Wa Municipal -2.285030 10.03062    258 274.7375   80
## 15  100105Home Economics Wa Municipal -2.285030 10.03062    242 258.1625   80
## 16     100106Agriculture Wa Municipal -2.285030 10.03062    223 240.6250   40
## 17        100106Business Wa Municipal -2.285030 10.03062    238 253.5000   40
## 18    100106General Arts Wa Municipal -2.285030 10.03062    248 268.9750   40
## 19        100201Business        Lawra -2.800941 10.54640    288 314.2750   80
## 20    100201General Arts        Lawra -2.800941 10.54640    319 339.0250   40
```

Exercise 3 Distance

```
##   HighSchool                          SeniorSchool StuId    jsslong    jsslat
## 1         NA Bosomtwe/Atwima/Kwanwoma (Kuntanase)     1 -1.5627517  6.559323
## 2         NA                          Ho Municipal     2  0.5261422  6.717607
## 3         NA                   Kwabre (Mamponteng)     3 -1.5414201  6.806778
## 4         NA           Kassena/Nankani (Navrongo)     4 -1.2174410 10.909423
## 5          NA             Atwima Mponua (Nyinahin)     5 -2.1771805  6.549507
## 6          NA                         Kumasi Metro     6 -1.5971872  6.682060
## 7          NA             Nanumba North (Bimbilla)     7 -0.1417642  8.816774
## 8          NA                 Jomoro (Half Assini)     8 -2.8032203  5.069508
## 9          NA                     East Akim (Kibi)     9 -0.4543442  6.178558
## 10         NA            Ejura/Sekyedumase (Ejura)    10 -1.3679653  7.462874
## 11         NA               Sekyere West (Mampong)    11 -1.1800768  7.199565
## 12         NA           Kassena/Nankani (Navrongo)    12 -1.2174410 10.909423
## 13         NA                         Agona Swedru    13 -0.7552425  5.617353
## 14         NA               Tolon Kunbungu (Tolon)    14 -1.1097199  9.527246
## 15         NA                   Accra Metropolitan    15 -0.1971153  5.607396
## 16         NA          Mpohor-Wassa East (Daboase)    16 -1.6975694  5.330796
## 17         NA            Ejura/Sekyedumase (Ejura)    17 -1.3679653  7.462874
## 18         NA                   Ga West (Amasaman)    18 -0.3975105  5.664688
## 19         NA            Wassa Amenfi (Asankragwa)    19 -2.3020179  5.725518
## 20         NA                                 Bole    20 -2.2666752  8.629696
##    ssslong ssslat JSdist
## 1       NA     NA     NA
## 2       NA     NA     NA
## 3       NA     NA     NA
## 4       NA     NA     NA
## 5       NA     NA     NA
## 6       NA     NA     NA
## 7       NA     NA     NA
## 8       NA     NA     NA
## 9       NA     NA     NA
## 10      NA     NA     NA
## 11      NA     NA     NA
## 12      NA     NA     NA
## 13      NA     NA     NA
## 14      NA     NA     NA
## 15      NA     NA     NA
## 16      NA     NA     NA
## 17      NA     NA     NA
## 18      NA     NA     NA
## 19      NA     NA     NA
## 20      NA     NA     NA
```

Exercise 4 Descriptive Characteristics

```
##   Rank Cutoff_mean Cutoff_sd Quality_mean Quality_sd Distance_mean Distance_sd
## 1    1    284.5812 59.705298     311.1536   52.96497      34.78904    52.23429
## 2    2    277.7861 51.430078     303.6828   44.73330      33.67427    47.81538
## 3    3    262.6396 43.985059     289.9210   37.49325      28.25628    42.75935
## 4    4    249.4498 38.069156     278.4302   31.91191      22.62548    38.28634
## 5    5    210.3753  8.185402     251.9085   12.88347      31.78886    29.18945
## 6    6    210.3297  8.582465     249.4862   11.20343      31.16226    28.54966
```

```
##   Interval Cutoff_mean Cutoff_sd Quality_mean Quality_sd Distance_mean
## 1    0-25%      223.77     17.45        245.6       7.55        193.69
## 2   25-50%      272.99      9.71       270.55       9.47        272.18
## 3   50-75%      306.67     11.59       307.82      11.61        305.79
## 4  75-100%      360.07     22.63        366.6         26        360.79
##   Distance_sd
## 1       25.23
## 2       11.01
## 3        11.7
## 4       19.75
```

Part 2

Exercise 5 Data creation

```
## ydum
##    0    1
## 4382 5618
```

Data have been created.

Exercise 6 OLS

```
## [1] "corr(Y,X1) = 0.21601496838707"
```

Now, the correlation between Y and X1 is quite different from 1.2, the designated coefficient. But the coefficient of X1 is very close to 1.2. Besides, if we standardize every variables (Y, X1, X2, X3), then the coefficient of new X1 (X1_1) will be similar to the correlation between Y and X1.

```
## [1] "after standardization: coeff(X1) = 0.205705404881561"
```

```
## [1] "coefficients:"
```

```
##          [,1]
##     2.4907098
## X1  1.1976226
## X2 -0.8970514
## X3  0.0875850
```

```
## [1] "standard errors:"
```

```
##                             X1            X2            X3
##     1.649836e-03 -6.048035e-04 -5.043875e-05 -1.439837e-04
## X1 -6.048035e-04  3.012891e-04  5.557060e-07  1.288086e-06
## X2 -5.043875e-05  5.557060e-07  8.273992e-06 -4.787112e-08
## X3 -1.439837e-04  1.288086e-06 -4.787112e-08  4.706056e-04
```

Exercise 7 Discrete choice

```
## initial  value 24689.277344
## iter   2 value 4289.017097
## iter   3 value 4251.184984
## iter   4 value 4209.797985
## iter   5 value 4198.168216
## iter   6 value 3021.365796
## iter   7 value 2521.019589
## iter   8 value 2303.352493
## iter   9 value 2229.544875
## iter  10 value 2214.628584
## iter  11 value 2214.605447
## iter  12 value 2213.755610
## iter  13 value 2213.464263
## iter  14 value 2213.463096
## iter  15 value 2213.334779
## iter  16 value 2213.313310
## iter  16 value 2213.313307
## iter  16 value 2213.313307
## final  value 2213.313307
## converged
```

```
## initial  value 12337.718958
## iter   2 value 3813.743197
## iter   3 value 3203.783208
## iter   4 value 3139.556932
## iter   5 value 3124.746731
## iter   6 value 2420.100956
## iter   7 value 2293.654430
## iter   8 value 2261.396203
## iter   9 value 2235.625304
## iter  10 value 2224.693119
## iter  11 value 2224.583551
## iter  12 value 2223.255132
## iter  13 value 2223.227459
## iter  14 value 2223.209703
## iter  15 value 2223.081465
## iter  16 value 2223.017353
## iter  16 value 2223.017344
## iter  16 value 2223.017344
## final  value 2223.017344
## converged
```

```
##       Probit:est  Probit:se Probit:t-value Probit:p-value   Logit:est
## cons  3.04275799 0.10007791     30.4038917  2.826563e-194  5.42656128
## X1    1.17235964 0.04292123     27.3142131  1.712261e-158  2.10060104
## X2   -0.90546589 0.01858996    -48.7072561   0.000000e+00 -1.61851270
## X3   -0.01124976 0.04647615     -0.2420544   8.087430e-01 -0.01963054
##        Logit:se Logit:t-value Logit:p-value
## cons 0.18557828    29.2413603 1.965090e-180
## X1   0.07936254    26.4684202 2.869053e-149
## X2   0.03670968   -44.0895371  0.000000e+00
## X3   0.08323300    -0.2358504  8.135536e-01
```

Both probit and logit model predict that higher X1 yields higher probability of being ydum=1 and higher X2 or X3 decrease the probability of being ydum=1. According to the p-value under both models, the coefficients of X1,X2 are all significant and the X3 is not significant.

Exercise 8 Marginal Effects

```
##      margin.probit.avg margin.probit.mean margin.logit.avg margin.logit.mean
## cons       1.709663378        1.495066090       3.04864284       2.657750966
## X1         0.658724865        0.576041589       1.18011794       1.028805195
## X2        -0.508762736       -0.444902736      -0.90928065      -0.792694205
## X3        -0.006321007       -0.005527593      -0.01102844      -0.009614392
```

Note: “avg” means evaluating average marginal effects in the sample; “mean” means evaluating marginal effect at the mean.