The PDF-version result output of First Assignment

Zekun Ge

(1) Exercise ONE

1. Number of students

We could see that there are 340823 student applicants.

2. Number of schools

There are 898 distinct schools.

3. Number of programs

There are 32 different programs.

4. Number of choices

There are 5546 different choices.

5. Number of missing test score

There are 179887 missing test score.

6. Number of applying to the same school

There are 5546 applicants who apply to the same school.

(2) Exercise TWO

# schoolcode	choice	ID	score	agey	male	choicepgm	jssdistrict	rankplace	rank
# 10101	10101Agriculture	230518	288	15	1	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230518	288	15	1	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230518	288	15	1	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230518	288	15	1	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230518	288	15	1	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230518	288	15	1	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230518	288	15	1	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230518	288	15	1	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230518	288	15	1	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230518	288	15	1	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230518	288	15	1	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230518	288	15	1	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230756	288	14	0	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230756	288	14	0	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230756	288	14	0	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230756	288	14	0	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230756	288	14	0	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230756	288	14	0	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230756	288	14	0	Agriculture	Accra Metropolitan	3	10101Agriculture
# 10101	10101Agriculture	230756	288	14	0	Agriculture	Accra Metropolitan	3	10101Agriculture

(the following part is attached to the right side of the above chart)

# cutoff_Low to Top	quality_Average Score	size_Number of Admission	X.y	schoolname
# 288	309.0889	45	3247	
# 288	309.0889	45	3560	
# 288	309.0889	45	6031	
# 288	309.0889	45	3881	
# 288	309.0889	45	408	EBENEZER SENIOR HIGH SCHOOL, DANSOMAN

#	288	309.0889	45	2881	
#	288	309.0889	45	328	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN
#	288	309.0889	45	453	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN
#	288	309.0889	45	590	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN
#	288	309.0889	45	552	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN
#	288	309.0889	45	3887	
#	288	309.0889	45	472	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN
#	288	309.0889	45	3247	
#	288	309.0889	45	3560	
#	288	309.0889	45	6031	
#	288	309.0889	45	3881	
#	288	309.0889	45	408	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN
#	288	309.0889	45	2881	
#	288	309.0889	45	328	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN
#	288	309.0889	45	453	EBENEZER SENIOR HIGH. SCHOOL, DANSOMAN

(the following part is attached to the right side of the above chart)

#	sssdistrict	ssslong	ssslat
#	Accra Metro	NA	NA
#	Accra Metro	NA	NA
#	Accra Metro	NA	NA
#	Accra Metro	NA	NA
#	Accra Metropolitan	-0.1971153	5.607396
#	Accra Metro	NA	NA
#	Accra Metropolitan	-0.1971153	5.607396
#	Accra Metropolitan	-0.1971153	5.607396
#	Accra Metropolitan	-0.1971153	5.607396
#	Accra Metropolitan	-0.1971153	5.607396
#	Accra Metro	NA	NA

```
#   Accra Metropolitan  -0.1971153   5.607396
#   Accra Metro           NA          NA
#   Accra Metro           NA          NA
#   Accra Metro           NA          NA
#   Accra Metro           NA          NA
#   Accra Metropolitan  -0.1971153   5.607396
#   Accra Metro           NA          NA
#   Accra Metropolitan  -0.1971153   5.607396
#   Accra Metropolitan  -0.1971153   5.607396
```

(3) Exercise Three

```
#                 sssdistrict    ssslong       ssslat     X   point_x     point_y       Dist
# 1029181       Kpando    0.2673851 6.896852   1 0.2076307 6.375762   36.27809
# 1029182       Kpando    0.2673851 6.896852   1 0.2076307 6.375762   36.27809
# 1029183     South Dayi        NA       NA         1 0.2076307 6.375762       NA
# 1029184    Ho Municipal 0.5261422 6.717607  1 0.2076307 6.375762   32.22676
# 1029185    Asuogyaman       NA       NA         1 0.2076307 6.375762       NA
# 1029186          Kpando        NA       NA         1 0.2076307 6.375762       NA
# 1029187       South Dayi        NA       NA        1 0.2076307 6.375762       NA
# 1029188          Kpando 0.2673851 6.896852   1 0.2076307 6.375762   36.27809
# 1029189    Ho Municipal       NA       NA        1 0.2076307 6.375762       NA
# 1029190          Kpando 0.2673851 6.896852   1 0.2076307 6.375762   36.27809
# 1029191       South Dayi        NA       NA        1 0.2076307 6.375762       NA
# 1029192           Kpando        NA       NA         1 0.2076307 6.375762       NA
# 1029193       South Dayi        NA       NA        1 0.2076307 6.375762       NA
# 1029194          Kpando 0.2673851 6.896852   1 0.2076307 6.375762   36.27809
# 1029195          Kpando 0.2673851 6.896852   1 0.2076307 6.375762   36.27809
# 1029196            Kpando        NA       NA   1 0.2076307 6.375762       NA
# 1029197        South Dayi        NA       NA   1 0.2076307 6.375762       NA
# 1029198          Kpando 0.2673851 6.896852   1 0.2076307 6.375762   36.27809
# 1029199        South Dayi        NA       NA   1 0.2076307 6.375762       NA
# 1029200            Kpando        NA       NA   1 0.2076307 6.375762       NA
```

(4) Exercise Four

Rank_The Lowest_To_Highest

1. rankplace == 1

   Cutoff_mean = 279.8704

   Cutoff_sd = 57.34419

   Quality_mean = 306.5762

   Quality_sd = 50.57871

   Size_mean = 93.04011

   Size_sd = 54.78786

2. rankplace == 2

   Cutoff_mean = 277.1518

   Cutoff_sd = 50.97056

   Quality_mean = 302.9698

   Quality_sd = 44.17802

   Size_mean = 93.68101

   Size_sd = 54.99501

3. rankplace == 3

   Cutoff_mean = 262.9438

   Cutoff_sd = 43.62898

   Quality_mean = 290.079

   Quality_sd = 37.06944

   Size_mean = 91.81311

   Size_sd = 55.59272

4. rankplace == 4

   Cutoff_mean = 250.1061

Cutoff_sd = 37.67753

Quality_mean = 278.8267

Quality_sd = 31.49667

Size_mean = 88.3362

Size_sd = 56.27731

5. rankplace == 5

   Cutoff_mean = 211.1775

   Cutoff_sd = 7.996356

   Quality_mean = 252.9305

   Quality_sd = 13.02076

   Size_mean = 68.32309

   Size_sd = 48.39169

6. rankplace == 6

   Cutoff_mean = 211.2391

   Cutoff_sd = 8.093442

   Quality_mean = 250.225

   Quality_sd = 11.02465

   Size_mean = 59.26986

   Size_sd = 46.24914

Score_The Lowest_To_Highest

1. Quantiles = 0.25

   Cutoff_mean = 217.7775

   Cutoff_sd = 14.31949

   Quality_mean = 252.3234

   Quality_sd = 12.39697

   Size_mean = 74.88889

   Size_sd = 14.31949

2. Quantiles = 0.5

Cutoff_mean = 229.8652

Cutoff_sd = 22.6533

Quality_mean = 262.2342

Quality_sd = 17.99262

Size_mean = 82.09206

Size_sd = 55.48278

3. Quantiles = 0.75

Cutoff_mean = 244.708

Cutoff_sd = 32.35489

Quality_mean = 274.8775

Quality_sd = 26.4583

Size_mean = 87.61462

Size_sd = 56.53001

3. Quantiles = 1.00

Cutoff_mean = 266.5514

Cutoff_sd = 51.04901

Quality_mean = 294.2175

Quality_sd = 44.18152

Size_mean = 90.61314

Size_sd = 55.32259

(5) Exercise Five

See the detailed information at R code.

(6) Exercise Six

6.1

Here betay_x1 approaches 1.26, which is very close to 1.2.

6.3

<u>When coding by hand</u>:

```
head(beta_All)
#          [,1]
#      2.4907098
# x1  1.1976226
# x2 -0.8970514
# x3  0.0875850
```

<u>When using lm():</u>

```
All_model$coefficients
#  (Intercept)          x1          x2          x3
#  2.4907098   1.1976226  -0.8970514   0.0875850
```

The corresponding results are exactly the same.

6.4

```
sqrt(diag(vcov_beta_hat))
#                  x1            x2            x3
#  0.040620200   0.017358550   0.002876599   0.021694530
```

(7) Exercise Seven

7.1.1

Linear Probability Model by optimization

```
reg_OLS_OP$par
#      alpha       beta1       beta2        beta3       sigma
# 0.886603815  0.146120426 -0.102847063 -0.008958128  0.334971033
```

Linear Probability Model by lm()

```
unname(coef(reg_OLS))
# [1]  0.885823611  0.146193985 -0.102832042 -0.008053057
```

The corresponding results are exactly the same.

7.1.2

Probit Model by optimization

```
reg_probit_op$par
#  (Intercept)          x1                   x2                x3
# 3.04273899  1.17235283 -0.90546035 -0.01124978
```

Probit Model by glm()

```
unname(coef(reg_probit))
# [1]  3.04273897  1.17235282 -0.90546040 -0.01124978
```

The corresponding results are approximately the same.

7.1.3

Logit Model by optimization

```
summary(reg_logit_OP, order = "convcode")
#                       X.Intercept.        x1              x2                  x3
#     BFGS          5.4265418 2.1005953  -1.6185076   -0.01963006
```

Logit Model by glm()

```
coef(reg_logit)
#  (Intercept)          x1                  x2                x3
# 5.42654014   2.10059417 -1.61850702 -0.01963017
```

The corresponding results are approximately the same.

7.2

```
reg_OLS <- lm(ydum ~ 1 + x1 + x2 + x3)
coef(reg_OLS)
```

```
#   (Intercept)          x1            x2           x3
# 0.885823611    0.146193985  -0.102832042 -0.008053057


reg_probit <- glm(ydum ~ 1 + x1 + x2 + x3,family = binomial(link = "probit"))
coef(reg_probit)
# (Intercept)          x1           x2          x3
# 3.04273897    1.17235282 -0.90546040 -0.01124978


reg_logit <- glm(ydum ~ 1 + x1 + x2 + x3,family = binomial(link = "logit"))
coef(reg_logit)
#   (Intercept)         x1           x2          x3
#  5.42654014    2.10059417 -1.61850702 -0.01963017
```

#/ The coefficients between probit regression and logit model are quite similar

#/ However, the OLS coefficients are smaller than the corresponding logit and probit figures.

#/ But for all these three models, the coefficient of each individual variable in any model remains negative or positive simultaneously.

For simplicity, I will use the variable "x1" as an example.

#/ OLS Estimation

#/ The coefficients of "x1" means that, holding other variables constant, one unit increase of x1 will lead to approximately 14% increase of ydum.

#/ probit Estimation and logit Estimation

#/ The coefficients of "x1" means nothing in either probit model or logit model.

#/ It only tells that, holding other variables constant, a positive "x1" coefficient means that an increase in the "x1" predictor leads to an increase in the predicted "ydum" probability

#/ Significance

#/ In each individual regression model, the coefficients of "Interpret", "x1", "x2" are statistically significant at 0.1% significance level. However, the coefficient of "x3" in each regression remains statistical insignificant.

(8) Exercise Eight

8.1

marginal_effects_Estimation_probit(reg_probit)

probit marginal.effects

x1   0.14380827

x2   -0.11106954

x3   -0.00137997

marginal_effects_Estimation_logit(reg_logit)

logit marginal.effects

x1   0.147623207

x2   -0.113743625

x3   -0.001379547

Using Packages as a double check:

Marginal effect in the Probit Model

```
Marginal Effects:
        dF/dx  Std. Err.         z  P>|z|    
x1  0.1438083  0.0043633   32.9588 <2e-16 ***
x2 -0.1110695  0.0004427 -250.8900 <2e-16 ***
x3 -0.0013805  0.0057218   -0.2413 0.8093    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Marginal effect in the Logit Model

```
Marginal Effects:
        dF/dx  Std. Err.        z  P>|z|    
x1  0.1440309  0.0072635  19.8293 <2e-16 ***
x2 -0.1109758  0.0044587 -24.8900 <2e-16 ***
x3 -0.0013465  0.0057815  -0.2329 0.8158    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

8.2

For simplicity, I use “mfx” package and “margins” package simultaneously.

Standard error of the marginal effects in the Probit Model

summary(margins(reg_probit))

```
factor     AME     SE          z      p   lower   upper
     x1  0.1438 0.0044   32.6602 0.0000  0.1352  0.1524
     x2 -0.1111 0.0004 -247.7190 0.0000 -0.1119 -0.1102
     x3 -0.0014 0.0057   -0.2419 0.8089 -0.0126  0.0098
```

Standard error of the marginal effects in the Logit Model

summary(margins(reg_logit))

```
# factor     AME     SE          z       p     lower   upper
#   x1     0.1440  0.0044   32.6564   0.0000   0.1354  0.1527
#   x2    -0.1110  0.0004 -246.8795   0.0000  -0.1119 -0.1101
#   x3    -0.0013  0.0057   -0.2359   0.8135  -0.0125  0.0098
```