

Exercise 1 Missing Data

number of students

[1] 340823

number of school

[1] 641

number of programs

[1] 33

number of choices

[1] 3086

number of missing test score

[1] 179887

students who apply to same school

[1] 120071

number of students who apply to less than 6 choices

[1] 18954

Exercise 2 Data

	sp <chr>	schoolcode <int>	programcode <chr>	sssdistrict <chr>	ssslong <dbl>	ssslat <dbl>	cutoff <dbl>	quality <dbl>	size <int>
1	100101	100101		Wa Municipal	-2.28503	10.03062	NA	NA	NA
2	100101 General Arts	100101	General Arts	Wa Municipal	-2.28503	10.03062	198	244.3924	79
3	100101 Home Economics	100101	Home Economics	Wa Municipal	-2.28503	10.03062	199	229.4500	40
4	100101 Technical	100101	Technical	Wa Municipal	-2.28503	10.03062	201	235.1020	49
5	100102 Agriculture	100102	Agriculture	Wa Municipal	-2.28503	10.03062	273	292.5556	90
6	100102 Business	100102	Business	Wa Municipal	-2.28503	10.03062	283	303.3444	90
7	100102 General Arts	100102	General Arts	Wa Municipal	-2.28503	10.03062	291	311.1111	90
8	100102 General Science	100102	General Science	Wa Municipal	-2.28503	10.03062	273	298.4333	90
9	100102 Home Economics	100102	Home Economics	Wa Municipal	-2.28503	10.03062	262	278.8667	45
10	100102 Visual Arts	100102	Visual Arts	Wa Municipal	-2.28503	10.03062	250	275.2000	45

1-10 of 20 rows

Previous **1** 2 Next

	sp <chr>	schoolcode <int>	programcode <chr>	sssdistrict <chr>	ssslong <dbl>	ssslat <dbl>	cutoff <dbl>	quality <dbl>	size <int>
11	100103 Agriculture	100103	Agriculture	Wa Municipal	-2.28503	10.03062	NA	NA	NA
12	100103 Business	100103	Business	Wa Municipal	-2.28503	10.03062	NA	NA	NA
13	100103 General Arts	100103	General Arts	Wa Municipal	-2.28503	10.03062	NA	NA	NA
14	100104 Business	100104	Business	Wa Municipal	-2.28503	10.03062	NA	NA	NA
15	100104 General Arts	100104	General Arts	Wa Municipal	-2.28503	10.03062	319	337.4444	45
16	100104 General Science	100104	General Science	Wa Municipal	-2.28503	10.03062	313	334.0000	45
17	100104 Home Economics	100104	Home Economics	Wa Municipal	-2.28503	10.03062	282	309.3556	45
18	100105	100105		Wa Municipal	-2.28503	10.03062	NA	NA	NA
19	100105 Business	100105	Business	Wa Municipal	-2.28503	10.03062	251	268.0125	80
20	100105 General Arts	100105	General Arts	Wa Municipal	-2.28503	10.03062	258	274.7375	80

11-20 of 20 rows

Previous **1** **2** Next

Exercise 3 Distance

Example: `library(ggplot2)`

	ssdis <fctr>	jssdis <fctr>	distance <chr>
1	Accra Metropolitan	South Dayi (Kpeve)	59.9919865044814
2	Ga West (Amasaman)	South Dayi (Kpeve)	64.419456204133
3	Tema	South Dayi (Kpeve)	50.6156936613713
4	Dangme East (Ada)	South Dayi (Kpeve)	32.6339719947566
5	Dangme West (Dodowa)	South Dayi (Kpeve)	45.8446614334538
6	Ga East (Abokobi)	South Dayi (Kpeve)	54.7920897810607
7	New Juaben (Koforidua)	South Dayi (Kpeve)	39.2072281117157
8	Suhum/Krabo/Coaltar	South Dayi (Kpeve)	55.6049736817609
9	Akwapim South (Nsawam)	South Dayi (Kpeve)	50.1630800018662
10	Akwapim North (Akropong)	South Dayi (Kpeve)	37.7225740726313

1-10 of 20 rows

Previous 1 2 Next

	ssdis <fctr>	jssdis <fctr>	distance <chr>
11	West Akim (Asamankese)	South Dayi (Kpeve)	67.519365351996
12	Birim South (Akim Oda)	South Dayi (Kpeve)	88.2993948724899
13	Birim North (New Abirem)	South Dayi (Kpeve)	85.1675633406139
14	Asuogyaman (Senchi Ferry)	South Dayi (Kpeve)	15.2887554205168
15	Kwahu North (Donkorkrom)	South Dayi (Kpeve)	49.2951979936485
16	Kwahu South (Mpraeso)	South Dayi (Kpeve)	60.3593355374014
17	Kwahu West (Nkawkaw)	South Dayi (Kpeve)	70.1887923329991
18	Kwaebibirem (Kade)	South Dayi (Kpeve)	71.2056590422866
19	Fanteakwa (Begoro)	South Dayi (Kpeve)	38.9767192057032
20	East Akim (Kibi)	South Dayi (Kpeve)	47.5074867842705

11-20 of 20 rows

Previous 1 2 Next

Exercise 4 Descriptive Characteristics

```
> print(table1)
      avg_cutoff sd_cutoff avg_quality sd_quality avg_distance sd_distance
rank1    317.0717  52.01054    337.9208  46.83908    34.14635    47.71170
rank2    299.6604  48.84759    321.3895  43.17346    32.92892    45.79090
rank3    287.1835  47.11865    310.1029  41.13584    30.67485    43.79528
rank4    273.9045  45.89648    298.6397  39.60251    26.48110    41.63369
rank5    256.5197  31.89327    284.1636  25.75738    30.26231    28.38107
rank6    251.4231  31.62788    279.6407  25.65834    30.84805    28.37482
>
> table_25 = change_name(table_25)
> print(table_25)
      avg_cutoff sd_cutoff avg_quality sd_quality avg_distance sd_distance
rank1    283.4534  44.59166    306.7529  38.94246    28.27913    44.89946
rank2    270.0009  41.40760    294.3333  35.90342    28.69947    43.71274
rank3    261.2568  40.43917    286.6283  34.93599    27.67535    42.15750
rank4    251.4020  39.75034    278.2927  34.20236    25.34848    40.95437
rank5    246.8855  31.22731    274.3982  25.74618    29.32590    28.78110
rank6    242.4731  30.73168    270.5157  25.52473    29.82050    28.64225
>
> table_50 = change_name(table_50)
> print(table_50)
      avg_cutoff sd_cutoff avg_quality sd_quality avg_distance sd_distance
rank1    301.0173  45.03198    322.7784  39.25254    31.69949    48.20433
rank2    285.1532  42.37029    308.0621  36.63525    31.22761    46.57879
rank3    273.9051  41.35209    298.0751  35.49417    29.81820    45.01575
rank4    262.0434  40.85982    288.0557  34.66926    26.12894    42.53877
rank5    253.4241  31.44824    280.8732  25.36836    29.94637    28.51642
rank6    248.7229  31.12694    276.7619  25.21741    30.73509    28.56983
>
> table_75 = change_name(table_75)
> print(table_75)
      avg_cutoff sd_cutoff avg_quality sd_quality avg_distance sd_distance
rank1    323.0130  43.53460    342.9473  38.32348    34.86944    48.94770
rank2    304.1427  42.36073    325.3159  36.63066    33.46690    46.83600
rank3    290.5017  41.65390    313.0342  35.69395    31.35241    44.98074
rank4    276.2045  41.29130    300.7063  34.98959    26.97811    42.57398
rank5    260.0547  31.18633    287.4211  24.47648    30.78063    28.11440
rank6    254.6174  31.21378    282.5944  24.75947    31.30439    28.20151
>
> table_100 = change_name(table_100)
> print(table_100)
      avg_cutoff sd_cutoff avg_quality sd_quality avg_distance sd_distance
rank1    362.1617  38.09107    380.4723  34.85716    41.98427    47.69021
rank2    340.5253  38.24956    358.9266  33.94823    38.49917    45.44718
rank3    324.1210  39.78167    343.6262  34.51467    33.97069    42.73751
rank4    306.9100  41.69904    328.3540  35.64904    27.51189    40.39295
rank5    266.4038  30.28669    294.6718  22.71537    31.05815    28.04709
rank6    260.5411  30.53754    289.3772  23.10501    31.59482    28.02791
```

Exercise 5 Data Creation

```
> x1[1:20]
[1] 1.227407 2.244599 2.218549 2.246759 2.721831 2.280621 1.018992 1.465101 2.332168
[10] 2.028502 2.387183 2.089950 1.565467 2.846867 1.584632 2.674591 1.572447 1.533642
[19] 1.373446 1.464452
> x2[1:20]
[1] 2.7310964 12.0238082 3.3911078 3.4090453 6.0187106 0.7536921 7.7037087
[8] 12.2640644 6.7635886 6.7426512 8.1600422 4.0815624 2.3634361 4.3802726
[15] 6.6004558 2.6077004 7.4811872 5.9427934 4.4314386 8.2670895
> x3[1:20]
[1] 0 0 1 1 0 0 0 1 0 0 1 0 1 0 0 0 1 0 0 0
> e[1:20]
[1] 2.3883829 4.4928205 3.6174976 1.3691200 3.6450301 1.9789293 1.0404695 3.9154502
[9] 2.4766193 0.6540830 1.8954995 1.6514924 1.7552042 1.9645526 4.4308561 1.8464937
[17] 0.6312023 1.6108071 0.8735494 1.4582745
> Y[1:20]
[1] 1.9032843 -3.1350882 3.8277599 1.5970899 1.9943875 4.5373519 -4.1700785
[8] -4.7640866 -0.3120093 -2.4801004 -1.9839194 0.9860258 2.1066723 1.9385475
[15] 0.8920039 3.2090729 -3.6149303 -1.3973371 -0.9666107 -3.7247639
> ydum[1:20]
[1] 1 0 1 1 1 1 0 0 1 0 0 1 1 1 1 1 0 0 0 0
```

Exercise 6 OLS

The correlation between Y and X1 is 0.2094291, which is smaller than 1.2

Coefficients for each variable

X1	1.1921859
X2	-0.9046174
X3	0.1099006

Exercise 7 Discrete Choice

For probit

```
Call:
glm(formula = ydum ~ X1 + X2 + X3, family = binomial(link = "probit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.5045 -0.1050  0.0063  0.2472  3.7964

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.96541    0.10011  29.621  <2e-16 ***
X1           1.23433    0.04459  27.684  <2e-16 ***
X2          -0.91993    0.01887 -48.761  <2e-16 ***
X3           0.13548    0.04694   2.886   0.0039 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13718.6  on 9999  degrees of freedom
Residual deviance:  4322.9  on 9996  degrees of freedom
AIC: 4330.9

Number of Fisher Scoring iterations: 8
```

The coefficient of X1 is 1.23, which is significant.

The coefficient of X2 is -0.91, which is significant.

The coefficient of X3 is 0.13, which is not significant.

1.23 means that one unit change in X1 will cause 1.23 unit change in ydum.

-0.91 means that one unit change in X2 will cause -0.91 unit change in ydum.

0.13 means that one unit change in X3 will cause 0.13 unit change in ydum.

For linear

```
Call:
lm(formula = ydum ~ X1 + X2 + X3)

Residuals:
    Min       1Q   Median       3Q      Max
-0.88888 -0.27076  0.06129  0.25277  1.49656

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.8911107  0.0135209   65.906  <2e-16 ***
X1           0.1452104  0.0057740   25.149  <2e-16 ***
X2          -0.1048519  0.0009664 -108.497  <2e-16 ***
X3           0.0107807  0.0071896    1.499    0.134
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3315 on 9996 degrees of freedom
Multiple R-squared:  0.5543,    Adjusted R-squared:  0.5541
F-statistic: 4144 on 3 and 9996 DF,  p-value: < 2.2e-16
```

The coefficient of X1 is 0.14, which is significant.

The coefficient of X2 is -0.10, which is significant.

The coefficient of X3 is 0.01, which is not significant.

0.14 means that one unit change in X1 will cause 0.14 unit change in ydum.

-0.10 means that one unit change in X2 will cause -0.10 unit change in ydum.

0.01 means that one unit change in X3 will cause 0.01 unit change in ydum.

For logit

```
Call:
glm(formula = ydum ~ X1 + X2 + X3, family = binomial)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.2035  -0.1433   0.0352   0.2598   3.3820

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  5.32125    0.18728  28.413 < 2e-16 ***
X1           2.21906    0.08304  26.722 < 2e-16 ***
X2          -1.65298    0.03763 -43.924 < 2e-16 ***
X3           0.24392    0.08435   2.892  0.00383 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 13718.6  on 9999  degrees of freedom
Residual deviance:  4334.3  on 9996  degrees of freedom
AIC: 4342.3

Number of Fisher Scoring iterations: 7
```

The coefficient of X1 is 2.21, which is significant.

The coefficient of X2 is -1.65, which is significant.

The coefficient of X3 is 0.24, which is not significant.

2.21 means that one unit change in X1 will cause 2.21 unit change in ydum.

-1.65 means that one unit change in X2 will cause -1.65 unit change in ydum.

0.24 means that one unit change in X3 will cause 0.24 unit change in ydum.

Exercise 8 Marginal Effects

```
probitmfx(formula = ydum ~ X1 + X2 + X3, data = data, atmean = TRUE)
```

Marginal Effects:

	dF/dx	Std. Err.	z	P> z	
X1	0.4924149	0.0177859	27.6857	< 2e-16	***
X2	-0.3669907	0.0075174	-48.8187	< 2e-16	***
X3	0.0539974	0.0186723	2.8918	0.00383	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

dF/dx is for discrete change for the following variables:

```
[1] "X3"
```

glm.fit: fitted probabilities numerically 0 or 1 occurredCall:

```
logitmfx(formula = ydum ~ X1 + X2 + X3, data = data, atmean = TRUE)
```

Marginal Effects:

	dF/dx	Std. Err.	z	P> z	
X1	0.5547272	0.0207585	26.723	< 2e-16	***
X2	-0.4132169	0.0093871	-44.020	< 2e-16	***
X3	0.0608904	0.0209894	2.901	0.00372	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

dF/dx is for discrete change for the following variables:

```
[1] "X3"
```

For probit

	marginal effect	standard error
X1	0.4924149	0.0177859
X2	-0.3669907	0.0075174
X3	0.0539974	0.0186723

For logit

	marginal effect	standard error
X1	0.5547272	0.0207585
X2	-0.4132169	0.0093871
X3	0.0608904	0.0209894