

Econ 613 HW 3

3/23/2022

1. Exercise 1 Basic Statistics

```
```{r}
library(tidyverse)
library(tidyr)
library(dplyr)
library(data.table)

getwd()
setwd("/Users/liulu/Desktop/Data")

datj <- read.csv(file = 'datjss.csv')
dats <- read.csv(file = 'datsss.csv')
datstu <- read.csv(file = 'datstu_v2.csv')
dats[!(is.na(dats$schoolname) | dats$schoolname=="") | is.na(dats$sssdistrict) | dats$sssdistrict==""),]
```

```

(1) Number of Students, schools, programs

```
nrow(datstu)
length(unique(dats$schoolcode))

programs <- c(datstu$choicepgm1,
datstu$choicepgm2,datstu$choicepgm3,datstu$choicepgm4,datstu$choicepgm5,datstu$choicepgm6)
programs1 <- nrow(unique(data.frame(programs)))
programs1

[1] 340823
[1] 898
[1] 33
```

(2) Number of choices (school, program)

```
long_datstu <- datstu %>% pivot_longer("schoolcode1":"schoolcode6", names_to = "schoolcode", values_to = "code")
%>% pivot_longer("choicepgm1":"choicepgm6", names_to = "choicepgm", values_to = "program")
long_datstu <- long_datstu %>%
  mutate(schoolcode = recode(schoolcode,
    "schoolcode1" = "1",
    "schoolcode2" = "2",
    "schoolcode3" = "3",
    "schoolcode4" = "4",
    "schoolcode5" = "5",
    "schoolcode6" = "6"))
long_datstu <- long_datstu %>%
  mutate(choicepgm = recode(choicepgm,
    "choicepgm1" = "1",
    "choicepgm2" = "2",
    "choicepgm3" = "3",
    "choicepgm4" = "4",
    "choicepgm5" = "5",
    "choicepgm6" = "6"))

long_datstu1 <- subset(long_datstu,long_datstu$schoolcode==long_datstu$choicepgm) %>% drop_na(code)
long_datstu1[!long_datstu1$program == "",]
long_datstu2 <- long_datstu1[c('code','program')]
nrow(unique(long_datstu2))
```

```
[1] 3080
```

(3) Number of students applying to at least one senior high schools in the same district to home

```
same_district_home <- datstu %>% select(V1, contains("school"), jssdistrict) %>% pivot_longer(cols = schoolcode1:schoolcode6, names_to = "school", values_to = "schoolcode")
highsch_district <- dats[,c(3,4)]
sameplace <- left_join(same_district_home, highsch_district, by = "schoolcode") %>% mutate(jssdistrict == sssdistrict) %>%
  distinct(V1, .keep_all = TRUE)
names(sameplace)[names(sameplace) == 'jssdistrict == sssdistrict'] <- 'match'
count(sameplace, sameplace$match == TRUE)
```

| sameplace\$match == TRUE | n |
|--------------------------|--------|
| <lgI> | <int> |
| FALSE | 194142 |
| TRUE | 146579 |
| NA | 102 |

(4) Number of students each senior high school admitted

```
numofs <- long_datstu1 %>% select(V1,score,rankplace,schoolcode,code) %>%
  filter(!is.na(rankplace) & rankplace != 99)
as.numeric(unlist(numofs))
numofs$admitted <- ifelse(numofs$rankplace == numofs$schoolcode,1,0)
student_admit <- numofs %>% filter(admitted == 1) %>% select(code) %>% group_by(code) %>% dplyr::summarise(count = n())
student_admitted <- unique(student_admit)
```

(5) The cutoff of senior high schools (the lowest score to be admitted

```
# A tibble: 517 x 2
  code `min(score)`
  <int>      <int>
1 10101     284
2 10102     343
3 10103     316
4 10104     245
5 10105     260
6 10106     293
7 10107     281
8 10108     248
9 10109     257
10 10110    343
# ... with 507 more rows
```

(6) The quality of senior high schools (the average score of students admitted)

```
# A tibble: 517 × 2
  code `mean(score)` 
  <int>      <dbl>
1 10101       320.
2 10102       394.
3 10103       354.
4 10104       297.
5 10105       351.
6 10106       340.
7 10107       312.
8 10108       304.
9 10109       282.
10 10110      408.
# ... with 507 more rows
```

2. Exercise 2 Data

(1) the district where the school is located

```
school_program <- long_datstu4 %>% mutate(choice = paste0(code,program))
names(school_program)[names(school_program) == 'code'] <- 'schoolcode'
sdistrict <- dats[,3:6] %>% select(schoolcode, sssdistrict, ssslong, ssslatt)
district <- sdistrict %>% distinct(schoolcode, sssdistrict, ssslong, ssslatt, .keep_all = TRUE)
district <- na.omit(district)
schooldata <- school_program %>% left_join(district, by = "schoolcode")

names(cutoff)[names(cutoff) == 'code'] <- 'schoolcode'
schooldata <- schooldata %>% left_join(cutoff, by = "schoolcode")

names(quality)[names(quality) == 'code'] <- 'schoolcode'
schooldata <- schooldata %>% left_join(quality, by = "schoolcode")

names(student_admitted)[names(student_admitted) == 'code'] <- 'schoolcode'
schooldata <- schooldata %>% left_join(student_admitted, by = "schoolcode")
schooldata
```

| | schoolcode | program | choice | sssdistrict | ssslong | ssslat | min(score) | mean(score) | count |
|----|------------|----------------|---------------------|-----------------------------|-------------|----------|------------|-------------|-------|
| 1 | 50112 | Home Economics | 50112Home Economics | Kumasi Metro | -1.59718716 | 6.682060 | 293 | 325.1623 | 2994 |
| 2 | 50107 | General Arts | 50107General Arts | Kumasi Metro | -1.59718716 | 6.682060 | 350 | 383.3051 | 3264 |
| 3 | 50202 | Visual Arts | 50202Visual Arts | Atwima / Nwabiagya (Nkawie) | -1.80875707 | 6.681337 | 314 | 334.8950 | 3600 |
| 4 | 50202 | Visual Arts | 50202Visual Arts | Atwima / Nwabiagya (Nkawie) | -1.80875707 | 6.681337 | 314 | 334.8950 | 3600 |
| 5 | 50702 | Home Economics | 50702Home Economics | Kwabre (Mamponteng) | -1.54142010 | 6.806778 | 242 | 283.9383 | 3600 |
| 6 | 50901 | General Arts | 50901General Arts | Ejura/Sekyedumase (Ejura) | -1.36796534 | 7.462874 | 211 | 254.0703 | 1536 |
| 7 | 70102 | General Arts | 70102General Arts | Ho Municipal | 0.52614224 | 6.717607 | 300 | 357.8523 | 2640 |
| 8 | 70602 | Business | 70602Business | Kpando | 0.26738513 | 6.896852 | 277 | 328.9750 | 1920 |
| 9 | 70107 | General Arts | 70107General Arts | Ho Municipal | 0.52614224 | 6.717607 | 213 | 250.8621 | 522 |
| 10 | 70105 | General Arts | 70105General Arts | Ho Municipal | 0.52614224 | 6.717607 | 199 | 249.9818 | 1650 |
| 11 | 70605 | Home Economics | 70605Home Economics | Kpando | 0.26738513 | 6.896852 | 202 | 253.2887 | 1746 |
| 12 | 70603 | General Arts | 70603General Arts | Kpando | 0.26738513 | 6.896852 | 210 | 274.0028 | 2160 |
| 13 | 50702 | Business | 50702Business | Kwabre (Mamponteng) | -1.54142010 | 6.806778 | 242 | 283.9383 | 3600 |
| 14 | 50705 | Home Economics | 50705Home Economics | Kwabre (Mamponteng) | -1.54142010 | 6.806778 | 218 | 262.1387 | 2292 |
| 15 | 50115 | Business | 50115Business | Kumasi Metro | -1.59718716 | 6.682060 | 298 | 327.7510 | 3060 |

3. Exercise 3 Distance

```
dist <- same_district_home %>% left_join(datj, by = "jssdistrict")
dats1 <- dats %>% select(schoolcode, sssdistrict, ssslong, ssslat)
dist1 <- dist %>% left_join(dats1, by = "schoolcode")
dist2 <- na.omit(dist1) %>% distinct(schoolcode, sssdistrict, ssslong, ssslat, .keep_all = TRUE)
dist2 %>% mutate(distance = sqrt((69.172 * (ssslong-point_x)*cos(point_y/57.3))^2 + (69.172*(ssslat - point_y))^2))
dist2
```

| | V1 | jssdistrict | school | schoolcode | X | point_x | point_y | sssdistrict | ssslong | ssslat | distance |
|----|----|--------------------------------------|-------------|------------|-----|------------|----------|-----------------------------|-------------|----------|-----------|
| 1 | 1 | Bosomtwe/Atwima/Kwanwoma (Kuntanase) | schoolcode2 | 50107 | 23 | -1.5627517 | 6.559323 | Kumasi Metro | -1.59718716 | 6.682060 | 8.813579 |
| 2 | 1 | Bosomtwe/Atwima/Kwanwoma (Kuntanase) | schoolcode1 | 50112 | 23 | -1.5627517 | 6.559323 | Kumasi Metro | -1.59718716 | 6.682060 | 8.813579 |
| 3 | 1 | Bosomtwe/Atwima/Kwanwoma (Kuntanase) | schoolcode3 | 50202 | 23 | -1.5627517 | 6.559323 | Atwima / Nwabiagya (Nkawie) | -1.80875707 | 6.681337 | 18.895053 |
| 4 | 1 | Bosomtwe/Atwima/Kwanwoma (Kuntanase) | schoolcode5 | 50702 | 23 | -1.5627517 | 6.559323 | Kwabre (Mamponteng) | -1.54142010 | 6.806778 | 17.179653 |
| 5 | 1 | Bosomtwe/Atwima/Kwanwoma (Kuntanase) | schoolcode6 | 50901 | 23 | -1.5627517 | 6.559323 | Ejura/Sekyedumase (Ejura) | -1.36796534 | 7.462874 | 63.917746 |
| 6 | 2 | Ho Municipal | schoolcode1 | 70102 | 117 | 0.5261422 | 6.717607 | Ho Municipal | 0.52614224 | 6.717607 | 0.000000 |
| 7 | 2 | Ho Municipal | schoolcode4 | 70105 | 117 | 0.5261422 | 6.717607 | Ho Municipal | 0.52614224 | 6.717607 | 0.000000 |
| 8 | 2 | Ho Municipal | schoolcode3 | 70107 | 117 | 0.5261422 | 6.717607 | Ho Municipal | 0.52614224 | 6.717607 | 0.000000 |
| 9 | 2 | Ho Municipal | schoolcode2 | 70602 | 117 | 0.5261422 | 6.717607 | Kpando | 0.26738513 | 6.896852 | 21.672792 |
| 10 | 2 | Ho Municipal | schoolcode6 | 70603 | 117 | 0.5261422 | 6.717607 | Kpando | 0.26738513 | 6.896852 | 21.672792 |
| 11 | 2 | Ho Municipal | schoolcode5 | 70605 | 117 | 0.5261422 | 6.717607 | Kpando | 0.26738513 | 6.896852 | 21.672792 |
| 12 | 3 | Kwabre (Mamponteng) | schoolcode3 | 50115 | 27 | -1.5414201 | 6.806778 | Kumasi Metro | -1.59718716 | 6.682060 | 9.439135 |
| 13 | 3 | Kwabre (Mamponteng) | schoolcode6 | 50703 | 27 | -1.5414201 | 6.806778 | Kwabre (Mamponteng) | -1.54142010 | 6.806778 | 0.000000 |
| 14 | 3 | Kwabre (Mamponteng) | schoolcode2 | 50705 | 27 | -1.5414201 | 6.806778 | Kwabre (Mamponteng) | -1.54142010 | 6.806778 | 0.000000 |
| 15 | 3 | Kwabre (Mamponteng) | schoolcode4 | 50706 | 27 | -1.5414201 | 6.806778 | Kwabre (Mamponteng) | -1.54142010 | 6.806778 | 0.000000 |

4. Exercise 4 Dimensionality Reduction

(1) Recode the schoolcode into its first three digits (*substr*). Call this new variable **scode rev**.

```

newdata <- na.omit(long_datstu)
newdata[!apply(newdata == "", 1, all), ]
newdata1 <- newdata %>% filter(!is.na(rankplace) & rankplace != 99) %>% mutate(scode_rev = substr(code, 1, 3))

```

| | V1 | score | agey | male | jssdistrict | rankplace | schoolcode | code | choicepgm | program | scode_rev |
|----|--------|-------|------|------|--------------|-----------|------------|-------|-----------|-----------------|-----------|
| 1 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 1 | 30905 | 1 | General Science | 309 |
| 2 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 1 | 30905 | 2 | General Arts | 309 |
| 3 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 1 | 30905 | 3 | General Arts | 309 |
| 4 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 1 | 30905 | 4 | Home Economics | 309 |
| 5 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 1 | 30905 | 5 | General Arts | 309 |
| 6 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 1 | 30905 | 6 | General Science | 309 |
| 7 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 2 | 30902 | 1 | General Science | 309 |
| 8 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 2 | 30902 | 2 | General Arts | 309 |
| 9 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 2 | 30902 | 3 | General Arts | 309 |
| 10 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 2 | 30902 | 4 | Home Economics | 309 |
| 11 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 2 | 30902 | 5 | General Arts | 309 |
| 12 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 2 | 30902 | 6 | General Science | 309 |
| 13 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 3 | 30902 | 1 | General Science | 309 |
| 14 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 3 | 30902 | 2 | General Arts | 309 |
| 15 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 3 | 30902 | 3 | General Arts | 309 |

(2) Recode the program variable into 4 categories: arts (general arts and visual arts), economics (business and home economics), science (general science) and others. Call this new variable **pgm_rev**.

| newdata1 %>% mutate(program = case_when(program == "General Arts" ~ "arts", program == "Visual Arts" ~ "arts", program == "Business" ~ "economics", program == "Home Economics" ~ "economics", program == "General Science" ~ "science", TRUE ~ "others")) | | | | | | | | | | | |
|--|--------|-------|------|------|--------------|-----------|------------|-------|-----------|-----------|-----------|
| names(newdata1)[names(newdata1) == 'program'] <- 'pgm_rev' | | | | | | | | | | | |
| | V1 | score | agey | male | jssdistrict | rankplace | schoolcode | code | choicepgm | pgm_rev | scode_rev |
| 1 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 1 | 30905 | 1 | science | 309 |
| 2 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 1 | 30905 | 2 | arts | 309 |
| 3 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 1 | 30905 | 3 | arts | 309 |
| 4 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 1 | 30905 | 4 | economics | 309 |
| 5 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 1 | 30905 | 5 | arts | 309 |
| 6 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 1 | 30905 | 6 | science | 309 |
| 7 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 2 | 30902 | 1 | science | 309 |
| 8 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 2 | 30902 | 2 | arts | 309 |
| 9 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 2 | 30902 | 3 | arts | 309 |
| 10 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 2 | 30902 | 4 | economics | 309 |
| 11 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 2 | 30902 | 5 | arts | 309 |
| 12 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 2 | 30902 | 6 | science | 309 |
| 13 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 3 | 30902 | 1 | science | 309 |
| 14 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 3 | 30902 | 2 | arts | 309 |
| 15 | 179888 | 249 | 16 | 0 | Agona Swedru | 5 | 3 | 30902 | 3 | arts | 309 |

(3) Create a new choice variable **choice_rev**.

```

      V1 score agey male jssdistrict rankplace schoolcode code choicepgm pgm_rev
      <int> <int> <int> <int> <chr> <int> <chr> <int> <chr> <chr>
1 179888 249 16 0 Agona Swedru 5 1 30905 1 General Science 309 309General Science
2 179888 249 16 0 Agona Swedru 5 1 30905 2 General Arts 309 309General Arts
3 179888 249 16 0 Agona Swedru 5 1 30905 3 General Arts 309 309General Arts
4 179888 249 16 0 Agona Swedru 5 1 30905 4 Home Economics 309 309Home Economics
5 179888 249 16 0 Agona Swedru 5 1 30905 5 General Arts 309 309General Arts
6 179888 249 16 0 Agona Swedru 5 1 30905 6 General Science 309 309General Science
7 179888 249 16 0 Agona Swedru 5 2 30902 1 General Science 309 309General Science
8 179888 249 16 0 Agona Swedru 5 2 30902 2 General Arts 309 309General Arts
9 179888 249 16 0 Agona Swedru 5 2 30902 3 General Arts 309 309General Arts
10 179888 249 16 0 Agona Swedru 5 2 30902 4 Home Economics 309 309Home Economics

```

(4) Recalculate the cutoff and the quality for each recoded choice.

```

newdata2$admitted <- ifelse(newdata2$rankplace == newdata2$schoolcode, 1, 0)
newdata2_admitted <- newdata2 %>% filter(admitted == 1)
new_cutoff <- newdata2_admitted %>% group_by(choice_rev) %>% summarise(min(score))
new_cutoff
new_quality <- newdata2_admitted %>% group_by(choice_rev) %>% summarise(mean(score))
new_quality

```

| | choice_rev | `min(score)` | | choice_rev | `mean(score)` |
|----|--------------|--------------|----|--------------|---------------|
| | <chr> | <int> | | <chr> | <dbl> |
| 1 | 100arts | 191 | 1 | 100arts | 272. |
| 2 | 100economics | 191 | 2 | 100economics | 265. |
| 3 | 100others | 191 | 3 | 100others | 259. |
| 4 | 100science | 204 | 4 | 100science | 293. |
| 5 | 101arts | 203 | 5 | 101arts | 337. |
| 6 | 101economics | 203 | 6 | 101economics | 332. |
| 7 | 101others | 203 | 7 | 101others | 325. |
| 8 | 101science | 203 | 8 | 101science | 365. |
| 9 | 102arts | 206 | 9 | 102arts | 314. |
| 10 | 102economics | 206 | 10 | 102economics | 311. |

(5) Consider the 20,000 highest score students.

```

highest <- newdata2 %>% arrange(desc(score)) %>% slice(1:2000)

```

| ▲ | V1 | score | agey | male | jssdistrict | rankplace | schoolcode | code | choicepgm | pgm_rev | scode_rev | choice_rev | admitted |
|----|--------|-------|------|------|--------------------|-----------|------------|-----------|-----------|--------------|-----------|------------|----------|
| 1 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 1 | 30107 1 | science | 301 | 301science | 1 | | |
| 2 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 1 | 30107 2 | economics | 301 | 301economics | 1 | | |
| 3 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 1 | 30107 3 | arts | 301 | 301arts | 1 | | |
| 4 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 1 | 30107 4 | economics | 301 | 301economics | 1 | | |
| 5 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 1 | 30107 5 | arts | 301 | 301arts | 1 | | |
| 6 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 1 | 30107 6 | arts | 301 | 301arts | 1 | | |
| 7 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 2 | 30107 1 | science | 301 | 301science | 0 | | |
| 8 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 2 | 30107 2 | economics | 301 | 301economics | 0 | | |
| 9 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 2 | 30107 3 | arts | 301 | 301arts | 0 | | |
| 10 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 2 | 30107 4 | economics | 301 | 301economics | 0 | | |
| 11 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 2 | 30107 5 | arts | 301 | 301arts | 0 | | |
| 12 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 2 | 30107 6 | arts | 301 | 301arts | 0 | | |
| 13 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 3 | 50102 1 | science | 501 | 501science | 0 | | |
| 14 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 3 | 50102 2 | economics | 501 | 501economics | 0 | | |
| 15 | 335624 | 469 | 15 | 0 | Accra Metropolitan | 1 3 | 50102 3 | arts | 501 | 501arts | 0 | | |