Homework 5

Xinyi Lin

10/12/2019

Problem 1

Quetion a

The estimated population mean score of life satisfaction is 5.812.

Question b

```
##             2.5 %   97.5 %
## STFLIFE 5.644571 5.979539
```

The confidence interval is (5.644571, 5.979539).

Question c

```
##             2.5 %   97.5 %
## STFLIFE 5.644149 5.979961
```

The 95% confidence interval is (5.644149, 5.979961).

Problem 2

Question a

```
##         mean     SE
## api00 656.59 9.2497
```

The mean of API score in 2000(api00) is 656.59 and standard error is 9.2497.

Question b

```
##         mean     SE
## api00 662.29 9.4089
```

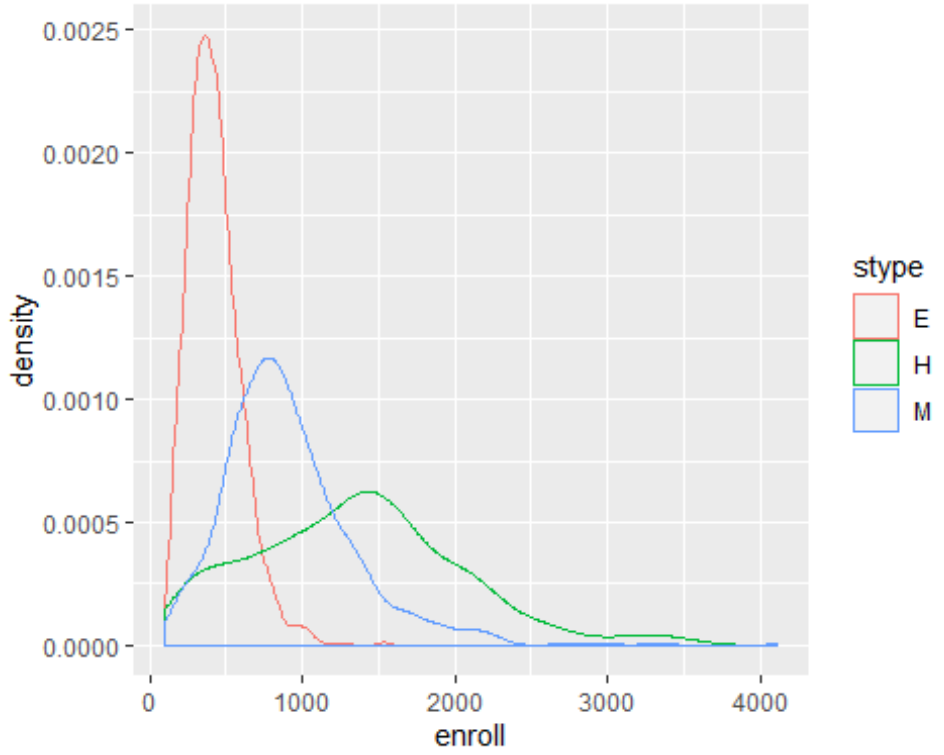
The mean of API score in 2000(api00) is 662.29 and standard error is 9.4089.

Question c

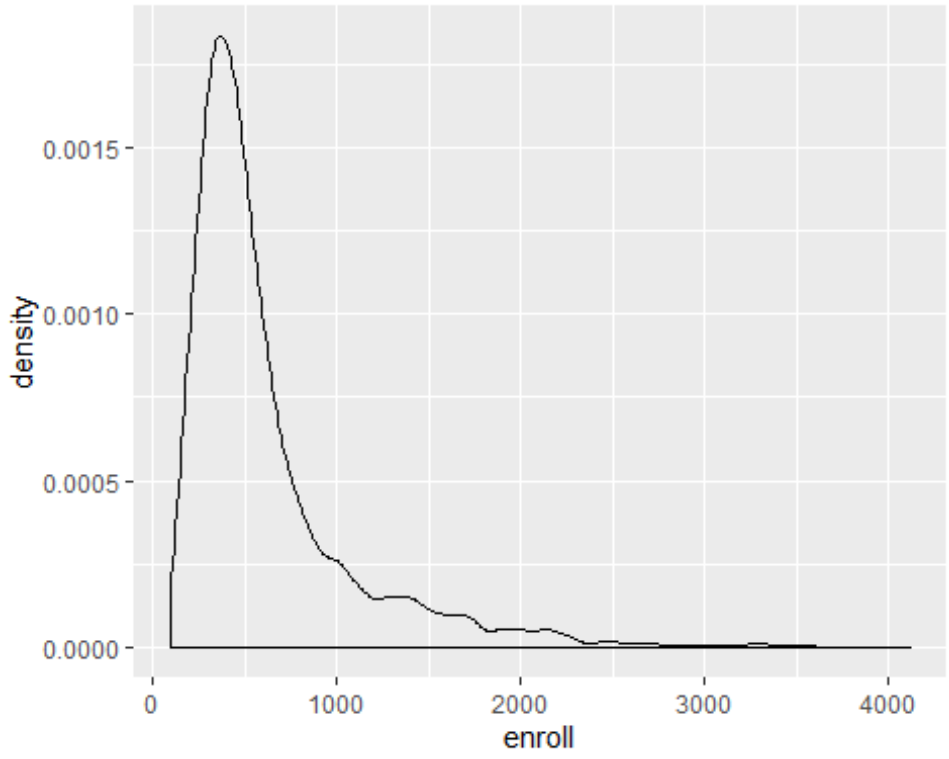
As the standard error obtained from stratification data 'apistrat' is slightly larger than that obtained from simple random sample data `apisrs`, there isn't a gain in precision by using the strtified design on estimating the mean API score.

Question d

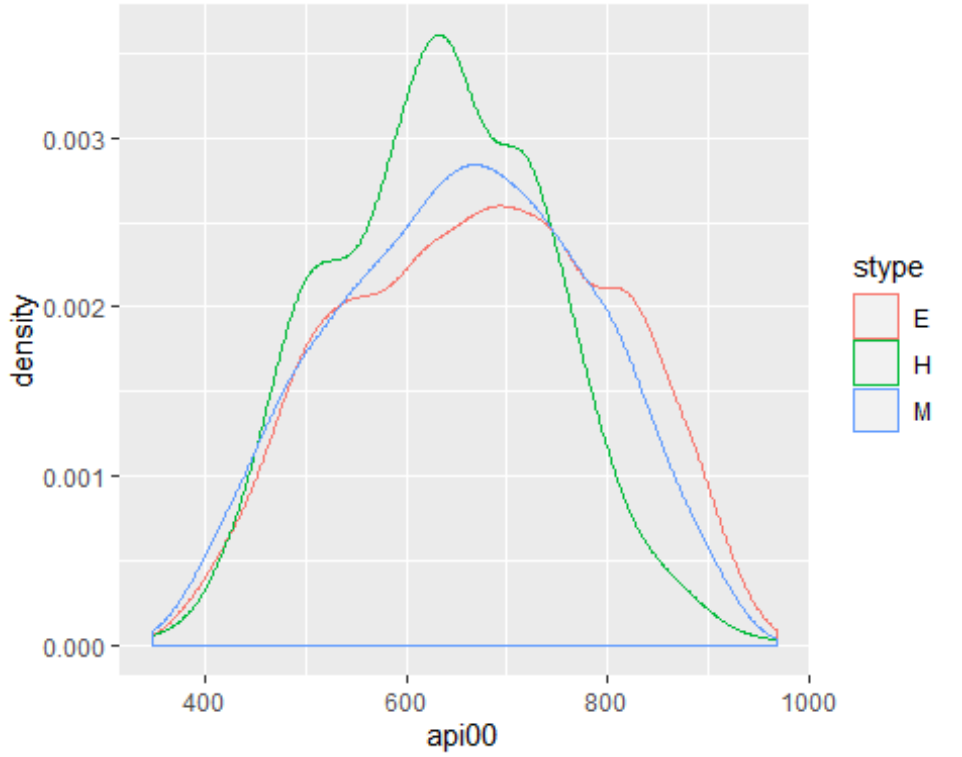
For enroll in three types of schools:



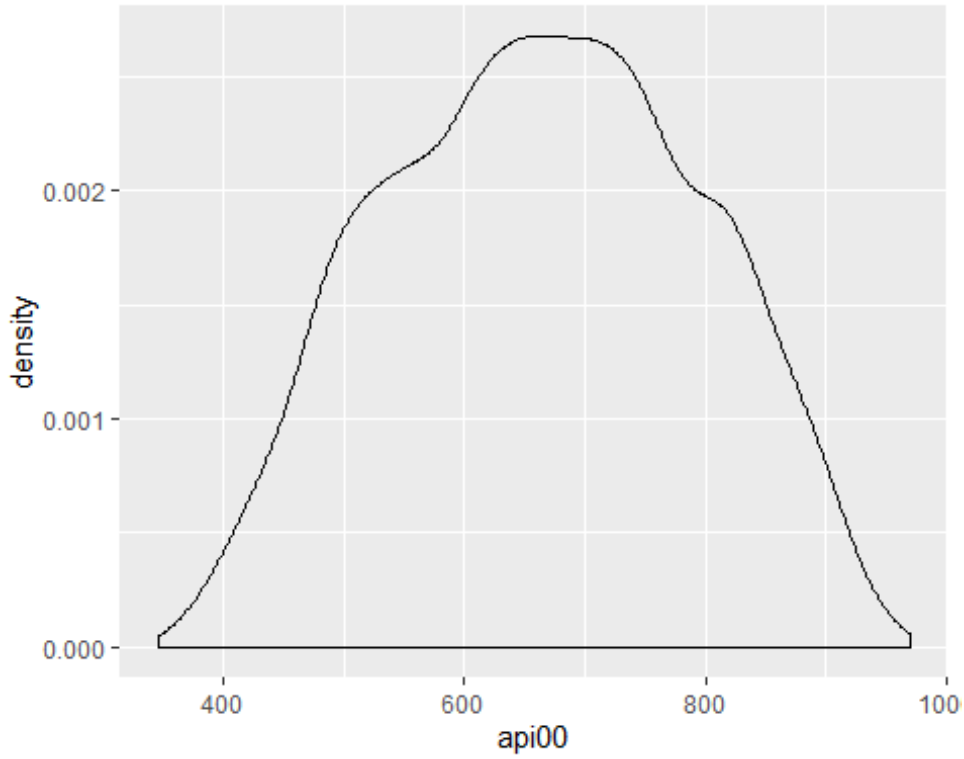
For enroll among all schools:



For API score in three types of schools:



For API score among all schools:



By comparing distributions of enroll and API score in three types of schools, we can find that distributions of enroll in three types of schools are significant different and distribution of enroll among all schools is also different from thoses in three types of schools. While distributions of API score in three types of schools and among all school are similar.

So if distributions of one survey outcome in different stratum are significant different and are different from the overall distribution, stratification can help to gain efficiency for estimating a population mean. But if distributions of one survey outcome in different stratum and among all dataset are similar, stratification cannot help to gain efficiency for estimating a population mean.

Appendix

```
knitr::opts_chunk$set(echo = FALSE)
library(survey)
library(tidyverse)
library(ggplot2)
data1 = read.csv("./ess6.csv") %>%
  mutate(psu = as.factor(psu))
q1_design = svydesign(id = ~psu, strata = ~stratify, weights = ~PSPWGHT, data
= data1)
q1_mean = svymean(~STFLIFE, q1_design, na.rm = TRUE)
confint(q1_mean)
JKn_design = as.svrepdesign(q1_design, type = 'JKn')
JKn_mean = svymean(~STFLIFE, JKn_design, na.rm = TRUE)
confint(JKn_mean)
data(api)
srs_design = svydesign(id=~1, fpc=~fpc, data = apisrs)
svymean(~api00, srs_design)
strat_design = svydesign(id=~1, strata = ~stype, fpc = ~fpc, data = apistrat)
svymean(~api00, strat_design)
apipop %>%
  ggplot(aes(x = enroll)) +
  geom_density(aes(color = stype))
apipop %>%
  ggplot(aes(x = enroll)) +
  geom_density()
apipop %>%
  ggplot(aes(x = api00)) +
  geom_density(aes(color = stype))
apipop %>%
  ggplot(aes(x = api00)) +
  geom_density()
```