POLI706: Advanced Methods of Political Analysis

Problem set 4

Exercise 1

For these analysis, you will be using the following dataset:

```
library(tidyverse);library(vdemdata)
vdemdata::vdem -> vdem

vdem |>
  dplyr::select(COWcode, year, v2x_polyarchy, e_gdppc) |>
  dplyr::filter(year == 2015) -> vdem_sub
```

a. Set the seed at 1224, then using a for-loop take a random sample of 5 countries' electoral democracy indices, 1,000 times. Save these averages. After than, set the seed at 1224, then using a for-loop take a random sample of 50 countries' electoral democracy indices 1,000 times. Save these averages.

b. Use a histogram to "look" at the distribution of averages we get with a sample size of 5 and a sample size of 50. How would you say they differ?

c. For the last set of averages, the ones obtained from a sample size of 50, what proportion are between 23 and 25?

Exercise 2

For these exercises, we will be using the following dataset:

```
library(tidyverse)
vdem |>
  dplyr::select(COWcode, year, e_boix_regime, e_gdppc, e_regionpol_6C) |>
  mutate(e_regionpol_6C =
           factor(e_regionpol_6C,
                  levels = c(1, 2, 3, 4, 5, 6),
                  labels = c("Eastern Europe and Central Asia",
                             "Latin America and the Caribbean",
                             "The Middle East and North Africa",
                             "Sub-Saharan Africa",
                             "Western Europe and North America",
```

```
                                "Asia and Pacific"))) |>
  drop_na() ->
  vdem_est
```

a. Consider `vdem_est` as the population. Use dplyr to create a vector x with the estimated GDP per capita of all democracies in the Sub-Saharan Africa. What is this population's average?

b. Compute the population standard deviation using the following formula:

$$\text{Population Standard Deviation} = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$$

c. Set the seed at 1234. Take a random sample X of size 25 from x. What is the sample average?

d. Consider `vdem_est` as the population. Use dplyr to create a vector y with the estimated GDP per capita of all democracies in the Asia and Pacific. What is this population's average?

e. Compute the population standard deviation using the following formula:

$$\text{Population Standard Deviation} = \sqrt{\frac{\sum(x-\bar{x})^2}{n}}$$

f. Set the seed at 1234. Take a random sample Y of size 25 from y. What is the sample average?

g. What is the difference in absolute value between $\bar{y}-\bar{x}$ and $\bar{Y}-\bar{X}$?

h. Repeat the above for autocracies. Make sure to set the seed to 1234 before each sample call. What is the difference in absolute value between $\bar{y}-\bar{x}$ and $\bar{Y}-\bar{X}$?

i. What is the difference between the sample averages?

Exercise 3

Simulate the promotion process for 24 men and 24 women. Assume 21 men and 14 women are recommended for promotion (assuming an overall promotion probability of 35/48 [total promoted/total]).

a. Calculate the difference in the proportion of men vs. women promoted, and repeat this simulation 10,000 times. Create a histogram of the differences in promotion proportions.

b. Count the number of occurrences where the difference in promotion proportion is at least 0.3 favoring men, and calculate the probability of witnessing at least a 0.3 difference.

c. Do you think the probability of gender discrimination in promotion is significant based on the results from `Exercise 3a` and `Exercise 3b`? Specifically, considering that the probability of observing a gender difference in promotion probability greater than 0.3, do you believe the initial observed sample indicating a large gender difference was an extreme case? Why or why not?