POLI706: Advanced Methods of Political Analysis

Problem set 2

Exercise 1

First, install the package, `tigerstats` and make an imaginary population using the following codes:

```
if (!require(tigerstats)) install.packages("tigerstats")
if (!require(manipulate)) install.packages("manipulate")

data(imagpop)     # load imaginary population data from tigerstats package
glimpse(imagpop)  # view the structure of data
help(imagpop)     # check the codebook
```

Let us say that the first 200 people in this population agree to be part of an experiment to see whether taking aspirin reduces the risk of heart disease.

```
AspHear <- imagpop[1:200, ]
```

The experiment involves randomly selecting 100 subjects to take an aspirin each morning for ten years, while the remaining 100 subjects, the control group, are given a placebo—a pill that looks and tastes like aspirin but has no effect on the body. The use of a placebo ensures that subjects remain unaware of their group assignment, preventing any knowledge that might influence lifestyle choices and affect the risk of heart disease. This approach is known as a single-blind experiment. Together, the aspirin and placebo groups are called the treatment groups because they receive different treatments. In this experiment, the X variable represents whether or not the subject takes aspirin, and the Y variable measures the subject's heart health. Since the X variable was assigned randomly, this design is referred to as a *completely randomized* design.
The R-function `RandomExp()` carries out the randomization:

```
Assignment <- RandomExp(AspHear,
                  sizes=c(100,100),
                  groups=c("placebo","aspirin"))
View(Assignment)
```

a. Do the treatment groups differ much with respect to **sex** (i.e., is **treat.grp** related to **sex**)?

b. Do the groups differ much with respect to **income**? (Is **treat.grp** related to **income**?)

c. Explain how properly conducted random assignment rules out selection bias.

d. Is the randomization procedure perfect to control potential confounding factors? Repeat the randomized experiments many times (more than 1,000 times) and draw the differences of **sex** and **income** between `placebo` group and `aspirin` group. From time to time, what do you discover? If the randomization procedure is perfect, what should you expect to observe?

Exercise2

a. Explain the relationship between *randomization* and *ceteris paribus*.

b. Explain the difference between *average treatment effect* and *individual treatment effect*.

Exercise 3

Simulate an experiment where 100 individuals are randomly assigned to either a treatment or a control group, and compare the baseline characteristics (`age` and `blood pressure`) of both groups to show that they are statistically similar before the treatment is applied. Use `rnorm()` function to compute hypothetical `age` and `blood_pressure` variables. Use `set.seed` for reproducibility.