Problem Set 3

Due date: 2 October

Please upload your completed assignment to the ELMs course site (under the assignments menu). Remember to include an annotated script file for all work with R and show your math for all other problems (if applicable, or necessary). Please also upload your completed assignment to the Github repository that you have shared with us. *We should be able to run your script with no errors.*

Total points: 30

```
#Disable warnings globally
options(warn=-1)
#.rs.restartR()

#Package loading
requiredPackages <- c("poliscidata",
                      "ggplot2",
                      "magrittr",
                      "dplyr",
                      "data.table",
                      "broom",
                      "rstatix",
                      "modelsummary")

#Installs packages if not yet installed
for (package in requiredPackages)
{
  if (!requireNamespace(package, quietly = TRUE))
  install.packages(package)

}
```

```
Registered S3 method overwritten by 'gdata':
```

```
  method         from
  reorder.factor gplots
```

```
invisible(lapply(requiredPackages, require, character.only = TRUE, quietly = T))
```

```
Attaching package: 'dplyr'

The following objects are masked from 'package:stats':

    filter, lag

The following objects are masked from 'package:base':

    intersect, setdiff, setequal, union


Attaching package: 'data.table'

The following objects are masked from 'package:dplyr':

    between, first, last


Attaching package: 'rstatix'

The following object is masked from 'package:stats':

    filter
```

```
rm(list=ls())
```

Question 1

Points: 10

Suppose I want to examine the hypothesis that consuming political news coverage from cable network channels (e.g., CNN, MSNBC, Fox News, etc.) causes citizens to have less trust in government than if they were to consume it from mainstream print media (e.g., New York Times, Wall Street Journal, etc.). What is more, suppose that I believe that this effect is greatest among "young" citizens. Design the contours of an experiment to test this proposition (and be sure to sketch the key elements of the experimental design). Second, briefly describe any significant shortcomings that the experimental design might have.Ans:

Ans:

This question has two parts, one to see the effects of network media vs print media in shaping government trust. The second part of the question wants see if this effect is the greatest amougst 'young' citizens.

Population definition:

I propose a stratified design, splitting participants into 'young' and 'mature' categories. With young being defined as 0-29 years of age and mature being 30+[1]. Participants can be recruited via online forms. We aim to get a population of about 80 participants equally split into these two categories.

Survey Questions:

1. What is your age - Options 1-100
2. What is your current trust in the Government - Scale of 1-10 where 1 indicates no trust
3. What percentage of your media consumption is between print media and network media - Sliding scale between 0 and 100% where 0 is totally print media while 100 is totally network media.

Experiment:

We can randomize the split between the categories to create our test and control groups. We can take about 20% of the block as the control group and allocate the rest as the experimental group [2]. We let the control group have the same media consumption as before. We can now split the remaining participants into two groups. One group will only consume print media and the other group will consume only network media for a period of one month [3]. At the end of this month, the participants can refill the survey. We then measure the statistical difference between the groups and see if the change in media watching was a significant effect.

Hypothesis:

We assume that the groups containing only network watchers will show a significant drop in trust compared to the control group. We should also see that there is no compareable significant drop in trust from those who watch print media.

[1] https://circle.tufts.edu/2022-election-center

[2] https://www.markhw.com/blog/control-size

[3] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6140520/

Shortcomings of the experiment:

1. We are not controlling for any other variables, other variables that affect trust in government such as political leaning, income level, educational levels, criminal background etc. will all have effects in how the government is perceived.
2. We should replicate this study many times in order to understand how strong the relationships are
3. Online forms will usually skew the sample to the younger participants.
4. Splitting the population into only two parts young and mature might hide a lot of variability in the mature population group. It can be split even further.
5. We don't really define what 'trust' means in the government as the citizens might trust their government in some areas but not in others. The questions needs to be defined very exactly.
6. There is no way of verifying if the participants are obeying the parameters of the experiment.

Question 2

Points: 10

Below are four hypothetical research designs to determine the effects of democratic governance on the funding of educational institutions. Briefly evaluate each sampling design in terms of its: (1) sampling variability; and, (2) sources of potential bias.

A. A sample of educational funding levels for every country in Europe for each of the last 100 years.

B. A sample of educational funding levels in five randomly picked countries drawn from the global population of countries for each of the last 10 years.

C. A survey of educational administrators that yields responses that allow for a comparison of educational funding levels in Russia and the United States for the six-year period from 2010 to 2015.

D. A sample of 75 randomly selected countries' educational funding levels for each of the past 25 years.

Ans:

A. 1. Sampling Variability: The sample is not randomized at at and draws from only a section of the world. It also does not address if a country was a democracy or not at the time of the observation. 2. Source of Bias: The effects of Europe cannot be generalized to the world. The recording of education will also be biased by not taking into account if the country was a democracy or not.

B. 1. Sampling Variability: The sample is not large enough to hold for the rest of the world. There are more than 195 countries in the world, a sample size of about 20% is needed to make any conclusion. It also does not take into account if the country was a democracy or not.

2. Source of Bias: The countries data will be skewed depending on the 5 countries picked. 10 years is also not a large enough time compared to most countries existence, as most countries are older than this.

C.

1. Sampling Variability: The sample only takes for two countries. Educational administrators may be a good source if it is across many countries.
2. Source of Bias: The sample size is too small. Also there is no definition of the country being a democracy or not, this skews results.The time period is also not large enough to ensure there is no bias.

D.

1. Sampling Variability: it has a good number of countries, but it has not taken into account if the country is a democracy or not.
2. Source of bias: There have been more than 2/3rds countries as democracies compared to non-democracies in the last 25 years. Pure randomization might not yield good results when one of the classes is so small.

Question 3

Points: 10

Using the `hanmerKalkanANES` dataset (posted on ELMs), construct a cross-tab with `presvote` (dichotomous variable coded with a value label) as the dependent variable and `ageCategorical` (1 = 18-30; 2 = 31-40; 3 = 41-50; 4 = 51-60; and 5 = 61-over) as the independent variable. Create a table with the cross-tab results and interpret them – i.e., what is the apparent association, if any, between age and U.S. presidential vote in 2004?

Next, what is:

A. The conditional distribution of respondents who are 41-50 that voted for Kerry?

presvote		1	2	3	4	5	All
Kerry	N	42	30	30	30	55	187
	% row	22.5	16.0	16.0	16.0	29.4	100.0
Bush	N	31	36	40	44	45	196
	% row	15.8	18.4	20.4	22.4	23.0	100.0
All	N	73	66	70	74	100	383
	% row	19.1	17.2	18.3	19.3	26.1	100.0

B. The conditional distribution of respondents who are 41-50 that voted for Bush?

```
library(dataverse)
library(foreign)

Sys.setenv("DATAVERSE_SERVER" = "dataverse.harvard.edu")


f <- get_file("hanmerKalkanANES.tab", "doi:10.7910/DVN/ARKOTI")

# load it into memory
tmp <- tempfile(fileext = ".dta")
writeBin(as.vector(f), tmp)
dat <- foreign::read.dta(tmp)
```

```
#Cross tab
dat <- drop_na(dat)
dat$age <- as.numeric(dat$age)

dat <- dat %>% mutate(ageCategorical =
                      case_when(age >= 18 & age <= 30 ~ 1,
                                age >= 31 & age <= 40 ~ 2,
                                age >= 41 & age <= 50 ~ 3,
                                age >= 51 & age <= 60 ~ 4,
                                age >= 61 ~ 5)
                                )

datasummary_crosstab(presvote ~ ageCategorical, data = dat)
```

Based on the crosstab, we can see that older people voted a little more in favor of bush compared to Kerry, together. but there was a strong showing for Kerry amongst the very old category.