Homework 3 – Intro to Probability and Statistics

Your name here

Instructions:

Due: 04/01 at 11:59PM.

What am I expecting? An R Markdown with the answers.

Have fun!

Changing Minds on Gay Marriage

In this assignment, you will analyze the data from two experiments in which households were canvassed for support on gay marriage.

This exercise is based on: LaCour, M. J., and D. P. Green. 2014. "When Contact Changes Minds: An Experiment on Transmission of Support for Gay Equality." *Science* 346(6215): 1366–69.

Context: Canvassers were given a script leading to conversations that averaged about twenty minutes. A distinctive feature of this study is that gay and straight canvassers were randomly assigned to households and canvassers revealed whether they were straight or gay in the course of the conversation. The experiment aims to test the 'contact hypothesis,' which contends that out-group hostility (towards gays in this case) diminishes when people from different groups interact with one another.

Name	Description
`study`	Study (1 or 2)
`treatment`	Treatment assignment: `No contact`, `Same-Sex Marriage Script by Gay Canvasser`, `Same-Sex Marriage Script by Straight Canvasser`, `Recycling Script by Gay Canvasser`, and `Recycling Script by Straight Canvasser`
`wave`	Survey wave (1-7). Note that Study 2 lacks wave 5 and 6.
`ssm`	Support for gay marriage (1 to 5). Higher scores indicate more support.

Each observation of this data set is a respondent giving a response to a five-point survey item on same-sex marriage. There are two different studies in this data set, involving interviews during 7 different time periods (i.e. 7 waves). In both studies, the first wave consists of the interview before the canvassing treatment occurs.

Use the gay.csv data set.

Question 1

Explore the dataset. Make histograms of the quantitative variables and barplots of the qualitative variables. Discuss them.

```
## Load tidyverse
library(tidyverse)
```

```
## -- Attaching packages --------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.0.9000     v purrr   0.3.3
## v tibble  2.1.3          v dplyr   0.8.3
```

```
## v tidyr   1.0.2          v stringr 1.4.0
## v readr   1.3.1          v forcats 0.4.0
```

```
## -- Conflicts ------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
## Loading the dataset
gay <- read.csv("https://raw.githubusercontent.com/umbertomig/intro-prob-stat-FGV/master/datasets/gay.cs

## Subsetting only study 1
gay <- filter(gay, study==1)

## Erasing the variable `study`
gay$study <- NULL

## Dataset head
head(gay, 10)
```

```
##     treatment wave ssm
## 1  No Contact    3   5
## 2  No Contact    4   5
## 3  No Contact    1   5
## 4  No Contact    6   5
## 5  No Contact    2   5
## 6  No Contact    7   5
## 7  No Contact    7   5
## 8  No Contact    6   4
## 9  No Contact    1   4
## 10 No Contact    2   4
```

```
## Your code here...
```

>> Your answer here <<

Question 2

Using the baseline interview wave before the treatment is administered (wave == 1), examine whether randomization was properly conducted. Base your analysis on the three groups:

- 'Same-Sex Marriage Script by Gay Canvasser'
- 'Same-Sex Marriage Script by Straight Canvasser'
- 'No Contact.'

Briefly comment on the results. Would you say that the study was well designed?

```
## Your code here...

## Hint 1. To subset a dataset, do the following:
## a. load the tidyverse package
## b. use the function 'filter(.)'
## c. if you don't know how to use the 'filter(.)', figure it out :)

## Hint 2.
## a. after the subset, you need to use the 'tapply(.)' function
## b. if you are having trouble using 'tapply(.)', see the (.Rmd) slides
```

>> Your answer here <<

Question 3

The second wave of survey was implemented two months after the canvassing. Estimate the average treatment effects (differences in means) of gay and straight canvassers on support for same-sex marriage, separately (wave == 2). Give a brief interpretation of the results.

```
## Hint: again, use filter(.) + tapply(.)

## Hint: compute the differences in means for gay versus none
# your code here...

## Hint: compute the differences in means for straight versus none
# your code here...

## Hint: I computed this in the minumum wage slide...
```

>> Your answer here <<

Question 4

Check if the results persist in wave 7 of the study. Give a brief interpretation of the results.

```
## Your code here...
```

>> Your answer here <<

Coronavirus

Rami Krispin build a dataset with coronavirus cases updated daily. The package name is `coronavirus`.

Question 1

Download the dataset and load it. Find the documentations for this package.

```
## Your code here...
```

>> Your answer here <<

Question 2

Can you find any stackoverflow resources about this package? If not, explain why? If yes, what types of resources?

>> Your answer here <<

Question 3

Make a table of number of cases by country. The table has to update itself automatically, so when I grade the problem set, the table will be the most updated one. Give a brief discussion on what did you do to solve this problem.

```
## Your code here...
```

>> Your answer here <<