# HW3_Markdown

2024-02-14


## Manay Divatia

## md46245


## 2/14/24

## Problem 1

### a

```
gay_df = read.csv("~/Downloads/gay.csv")
wave1_df = subset(gay_df, subset = wave == 1)
summary(wave1_df)
```

```
##      study        treatment              wave        ssm
##  Min.   :1.000   Length:11948       Min.   :1   Min.   :1.000
##  1st Qu.:1.000   Class :character   1st Qu.:1   1st Qu.:1.000
##  Median :1.000   Mode  :character   Median :1   Median :3.000
##  Mean   :1.204                      Mean   :1   Mean   :3.036
##  3rd Qu.:1.000                      3rd Qu.:1   3rd Qu.:5.000
##  Max.   :2.000                      Max.   :1   Max.   :5.000
```

```
no_contact_1 = subset(wave1_df, subset = (treatment == "No Contact"))
dim(no_contact_1)
```

```
## [1] 6441    4
```

```
marriage_gay_1 = subset(wave1_df, subset = (treatment == "Same-Sex Marriage
                                            Script by Gay Canvasser"))
dim(marriage_gay_1)
```

```
## [1] 0 4
```

```
marriage_straight_1 = subset(wave1_df, subset = (treatment == "Same-Sex Marriage
                                                 Script by Straight Canvasser"))
dim(marriage_straight_1)
```

```
## [1] 0 4
```

We see a lot more instances of no contact than with any other treatment which may suggest that it wasn't random after all. Another thing to note is that the treatment of Same-Sex Marriage Script by Gay Canvasser seemed to be almost double of the other treatments aside from the No Contact treatment. This also may indicate that the study was not random. ## b

```
mean(subset(gay_df, subset=(wave == 2 & study == 1 & treatment == "Same-Sex
                            Marriage Script by Gay Canvasser"))$ssm)
```

```
## [1] NaN
```

```
mean(subset(gay_df, subset=(wave == 2 & study == 1 & treatment == "Same-Sex
                            Marriage Script by Straight Canvasser"))$ssm)
```

```
## [1] NaN
```

From this, we can see that the average support for study 1 where there were gay canvassers was 3.13 and for that of straight canvassers it was 3.16. This is not really a large difference to indicate anything with full certainty. ## c

```
mean(subset(gay_df, subset=(wave == 2 & study == 1 & treatment == "Same-Sex
                            Marriage Script by Gay Canvasser"))$ssm)
```

```
## [1] NaN
```

```
mean(subset(gay_df, subset=(wave == 2 & study == 1 & treatment == "Recycling
                            Script by Gay Canvasser"))$ssm)
```

```
## [1] NaN
```

```
mean(subset(gay_df, subset=(wave == 2 & study == 1 & treatment == "Same-Sex
                            Marriage Script by Straight Canvasser"))$ssm)
```

```
## [1] NaN
```

```
mean(subset(gay_df, subset=(wave == 2 & study == 1 & treatment == "Recycling
                            Script by Straight Canvasser"))$ssm)
```

```
## [1] NaN
```

I think the purpose of having authors use a script to encourage people to recycle helped to show some inherent biases people may have even when hearing about nothing related to same sex marriage. When comaparing the results of the gay canvassers for the same-sex marriage and recycling scripts, we found that the same-sex marriage script had a 3.13 support while the recycling had 3.1 which isn't enough of a difference to make any claims. When comaparing the results of the straight canvassers for the same-sex marriage and recycling scripts, we found that the same-sex marriage script had a 3.16 support while the recycling had 3 which is slightly larger and may be indicative of some trend. ## d

```
mean(subset(gay_df, subset=(wave == 7 & study == 1 & treatment == "Same-Sex
                            Marriage Script by Gay Canvasser"))$ssm)
```

```
## [1] NaN
```

```
mean(subset(gay_df, subset=(wave == 7 & study == 1 & treatment == "Same-Sex
                            Marriage Script by Straight Canvasser"))$ssm)
```

```
## [1] NaN
```

It seems that there could be some lasting effects. After computing the mean, we find that with the gay canvasser, the average was 3.37 while with the straight canvasser, it was 3.27. This difference is about .1 which is something that could or couldn't be attributed to the canvassers. ## e

```
no_contact_2 = subset(gay_df, subset = (wave == 1 & study == 2 & treatment ==
                                          "No Contact"))
dim(no_contact_2)
```

```
## [1] 1203    4
```

```
marriage_gay_2 = subset(gay_df, subset = (wave == 1 & study == 2 &
                                            treatment == "Same-Sex Marriage
```

```
                                          Script by Gay Canvasser"))
dim(marriage_gay_2)
```

```
## [1] 0 4
```

It does seem that there is some randomization because the subsets seems to be around the same size. Probably what is more important is that there isn't a drastic difference. ## f

```
mean(subset(gay_df, subset=(wave == 2 & study == 2 & treatment == "Same-Sex
                            Marriage Script by Gay Canvasser"))$ssm)
```

```
## [1] NaN
```

Yes it looks to be consistent to some extent. The data showed in the first study a ssm of 3.13 while in the second study, a ssm of 3.11. If they were properly randomized, they wouldn't be too different in score. ## g

```
mean(subset(gay_df, subset=(wave == 1 & study == 2 & treatment == "Same-Sex
                            Marriage Script by Gay Canvasser"))$ssm)
```

```
## [1] NaN
```

```
mean(subset(gay_df, subset=(wave == 2 & study == 2 & treatment == "Same-Sex
                            Marriage Script by Gay Canvasser"))$ssm)
```

```
## [1] NaN
```

```
mean(subset(gay_df, subset=(wave == 3 & study == 2 & treatment == "Same-Sex
                            Marriage Script by Gay Canvasser"))$ssm)
```

```
## [1] NaN
```

```
mean(subset(gay_df, subset=(wave == 4 & study == 2 & treatment == "Same-Sex
                            Marriage Script by Gay Canvasser"))$ssm)
```

```
## [1] NaN
```

```
mean(subset(gay_df, subset=(wave == 7 & study == 2 & treatment == "Same-Sex
                            Marriage Script by Gay Canvasser"))$ssm)
```

```
## [1] NaN
```

The first wave is the lowest with a score of 2.97. The next 3 waves hover around 3.1 +- .1. The final wave jumps up to 3.33. Ultimately, it looks like the score increases throughout the waves in study 2. However, without the 5th or 6th wave, we can't know the full story.

# Problem 2

## a

```
leaders = read.csv("~/Downloads/leaders.csv")
summary(leaders)
```

```
##       year         country           leadername            age        
##  Min.   :1878   Length:250         Length:250         Min.   :18.00  
##  1st Qu.:1920   Class :character   Class :character   1st Qu.:45.00  
##  Median :1949   Mode  :character   Mode  :character   Median :52.50  
##  Mean   :1945                                         Mean   :53.52  
##  3rd Qu.:1972                                         3rd Qu.:61.75  
##  Max.   :2001                                         Max.   :81.00
```

```
##   politybefore      polityafter      interwarbefore  interwarafter
##  Min.   :-10.000   Min.   :-10.000   Min.   :0.000   Min.   :0.000
##  1st Qu.: -7.000   1st Qu.: -7.000   1st Qu.:0.000   1st Qu.:0.000
##  Median : -3.000   Median : -3.167   Median :0.000   Median :0.000
##  Mean   : -1.519   Mean   : -1.650   Mean   :0.188   Mean   :0.148
##  3rd Qu.:  4.000   3rd Qu.:  3.917   3rd Qu.:0.000   3rd Qu.:0.000
##  Max.   : 10.000   Max.   : 10.000   Max.   :1.000   Max.   :1.000
##  civilwarbefore  civilwarafter      result
##  Min.   :0.000   Min.   :0.000   Length:250
##  1st Qu.:0.000   1st Qu.:0.000   Class :character
##  Median :0.000   Median :0.000   Mode  :character
##  Mean   :0.216   Mean   :0.184
##  3rd Qu.:0.000   3rd Qu.:0.000
##  Max.   :1.000   Max.   :1.000
```

```
dim(leaders)
```

```
## [1] 250  11
```

```
unique(leaders$country)
```

```
##  [1] "Afghanistan"       "Albania"           "Algeria"
##  [4] "Argentina"         "Australia"         "Austria"
##  [7] "Belgium"           "Bhutan"            "Bolivia"
## [10] "Brazil"            "Burundi"           "Bulgaria"
## [13] "Cambodia"          "Canada"            "Ivory Coast"
## [16] "Chad"              "Chile"             "China"
## [19] "Colombia"          "Congo Brazzaville" "Costa Rica"
## [22] "Cuba"              "Cyprus"            "Czechoslovakia"
## [25] "Dominican Rep"     "Congo Kinshasa"    "Ecuador"
## [28] "Egypt"             "Ethiopia"          "France"
## [31] "Ghana"             "Germany"           "Greece"
## [34] "Georgia"           "Guatemala"         "Guinea"
## [37] "Haiti"             "Honduras"          "India"
## [40] "Indonesia"         "Iran"              "Iraq"
## [43] "Israel"            "Italy"             "Jordan"
## [46] "Japan"             "Kenya"             "Kuwait"
## [49] "Liberia"           "Lebanon"           "Libya"
## [52] "Madagascar"        "Mexico"            "Myanmar (Burma)"
## [55] "Nepal"             "Nicaragua"         "Niger"
## [58] "Netherlands"       "Oman"              "Pakistan"
## [61] "Panama"            "Peru"              "Poland"
## [64] "Portugal"          "Korea South"       "Russia"
## [67] "Vietnam South"     "Rwanda"            "South Africa"
## [70] "El Salvador"       "Saudi Arabia"      "Senegal"
## [73] "Somalia"           "Spain"             "Sri Lanka"
## [76] "Sudan"             "Sweden"            "Syria"
## [79] "Togo"              "Turkey"            "Uganda"
## [82] "United Kingdom"    "Uruguay"           "United States"
## [85] "Uzbekistan"        "Venezuela"         "Yemen North"
## [88] "Yugoslavia"
```

```
(2001 - 1878) / nrow(leaders)
```

```
## [1] 0.492
```

We can see from the dim() method that there are 250 assassination attempts. From unique(), we know that

there are 88 countries that have experience at least one leader assassination attempt. From the summary function, we found that this data ranged from 1878 to 2001. Finding mean assassinations per year, we get .492 which could be representated as about 1 assassination attempt every 2 years. ## b

```
leaders$dead <- ifelse(leaders$result == "dies within a day after the attack" |
                         leaders$result == "dies between a day and a week" |
                         leaders$result == "dies between a week and a month" |
                         leaders$result == "dies, timing unknown", 1, 0)
```

It does speak to the validity of the assumption that attempts are randomly determined because there seems to be a distribution of ways and counts in which a leader dies or doesn't die. ## c

```
mean(leaders$politybefore)
```

```
## [1] -1.518667
```

```
mean(leaders$polityafter)
```

```
## [1] -1.65
```

```
mean(leaders$age[leaders$dead == 0])
```

```
## [1] 52.71429
```

```
mean(leaders$age[leaders$dead == 1])
```

```
## [1] 56.46296
```

The polity before is -1.51 and after is -1.65. This shows a movement towards hereditary monarchy as defined by the Polity Project. As for age, the average age for leaders who survived was 52.7 while the average age for leaders who died was 56.5 which means that older leaders were maybe more likely to die. ## d

```
mean(leaders$interwarbefore)
```

```
## [1] 0.188
```

```
mean(leaders$interwarafter)
```

```
## [1] 0.148
```

```
mean(leaders$civilwarbefore)
```

```
## [1] 0.216
```

```
mean(leaders$civilwarafter)
```

```
## [1] 0.184
```

```
leaders$warbefore <- ifelse(leaders$civilwarbefore == 1 |
                              leaders$interwarbefore == 1, 1, 0)
mean(leaders$warbefore)
```

```
## [1] 0.368
```

In terms of international and civil war, both dropped in percentage from before to after. For international, it went from .188 to .148 and for civil war, it went from .216 to .184. After creating the warbefore column and looking at the mean, we find that 36.8% of assassinations had a civil or international war happen within 3 years prior to the assassination attempt.

**e**

```r
leader_dead_df = subset(leaders, subset=(dead == 1))
mean(leader_dead_df$polityafter - leader_dead_df$politybefore)
```

```
## [1] -0.05864198
```

```r
leaders$warafter <- ifelse(leaders$civilwarafter == 1 |
                             leaders$interwarafter == 1, 1, 0)
mean(leaders$warafter[leaders$dead == 1] - leaders$warbefore[leaders$dead == 1])
```

```
## [1] -0.1481481
```

From my findings, it looks like a successful leader assassination actually leads to more hereditary monarchy. This is because the average difference is -0.5. Since the figure is negative, it means that after assassinations, countries tend ot move towards monarchies. From my findings, it looks like that war decreases after a successful assasination. This is because my findings show that war after is lower than war before which means that war went down. Specifically, the mean was -.148 or -14.8%.