# Problem Set 1 Final Product

Nathaniel Williams

2025-10-13

## Simulation

### Setup

```
set.seed(12345)
Population <- 1000000
Role <- c('Runeblade','Guardian','Wizard','Ranger')
#Add in probablity weights to better show convergence effect
Role_probs<- c(0.4,0.3,0.2,0.1)
P_Roles <- sample(Role, Population, replace = TRUE, prob = Role_probs)
Groups <- sample(c("Treatment","Control"), Population, replace = TRUE)

## Df
Role_Data <- data.frame(
  ID = 1:Population,
  Role = P_Roles,
  Group = Groups
)
## parameters
n_values <- c(10, 25, 50, 100)
Repetitions <- 100
Overall_Results <- list()
```

### The Simulation

```
## sim
for (n in n_values) {
  Simulation_results <- list()

  for (i in 1:Repetitions) {

    #Rndm smple
    sample_indices <- sample(1:Population, n, replace = FALSE)
    sample_data <- Role_Data[sample_indices, ]

    #Ovll role prop
    overall <- prop.table(table(sample_data$Role))
    overall_df <- as.data.frame(overall)
    overall_df$GroupType <- "Overall"

    #Trtmnt grp prop
```

```
    treat_props <- prop.table(table(sample_data$Role[sample_data$Group == "Treatment"]))
    treat_df <- as.data.frame(treat_props)
    treat_df$GroupType <- "Treatment"

    #Cntrl grp prop
    control_props <- prop.table(table(sample_data$Role[sample_data$Group == "Control"]))
    control_df <- as.data.frame(control_props)
    control_df$GroupType <- "Control"

    #Combine
    df <- rbind(overall_df, treat_df, control_df)
    df$Iteration <- i
    df$SampleSize <- n

    # put in list
    Simulation_results[[i]] <- df
  }

  # Combine
  Overall_Results[[paste0("n_", n)]] <- do.call(rbind, Simulation_results)
}

## New df
simulation_df <- do.call(rbind, Overall_Results)

head(simulation_df)
```

```
##             Var1 Freq GroupType Iteration SampleSize
## n_10.1  Guardian  0.5   Overall         1         10
## n_10.2 Runeblade  0.3   Overall         1         10
## n_10.3    Wizard  0.2   Overall         1         10
## n_10.4  Guardian  0.6 Treatment         1         10
## n_10.5 Runeblade  0.2 Treatment         1         10
## n_10.6    Wizard  0.2 Treatment         1         10
```

## Data Visualzation

**As n increases we get closer to the population distributions**

```
pop_props <- prop.table(table(Role_Data$Role))

#Mean
avg_results <- aggregate(Freq~Var1 + GroupType + SampleSize,
                         data = simulation_df,
                         FUN = mean)

head(avg_results)
```

```
##        Var1 GroupType SampleSize      Freq
## 1  Guardian   Control         10 0.3622869
## 2 Runeblade   Control         10 0.4239418
## 3    Wizard   Control         10 0.3142507
## 4    Ranger   Control         10 0.2714866
## 5  Guardian   Overall         10 0.2909091
```

```
## 6 Runeblade   Overall         10 0.3930000
```

```
#Number table
smple_poptble <- with(avg_results[avg_results$GroupType=="Overall", ],
                      tapply(Freq, list(Var1, SampleSize), mean))
print(round(smple_poptble, 2))
```

```
##             10   25  50 100
## Guardian  0.29 0.31 0.3 0.3
## Runeblade 0.39 0.39 0.4 0.4
## Wizard    0.22 0.20 0.2 0.2
## Ranger    0.17 0.11 0.1 0.1
```

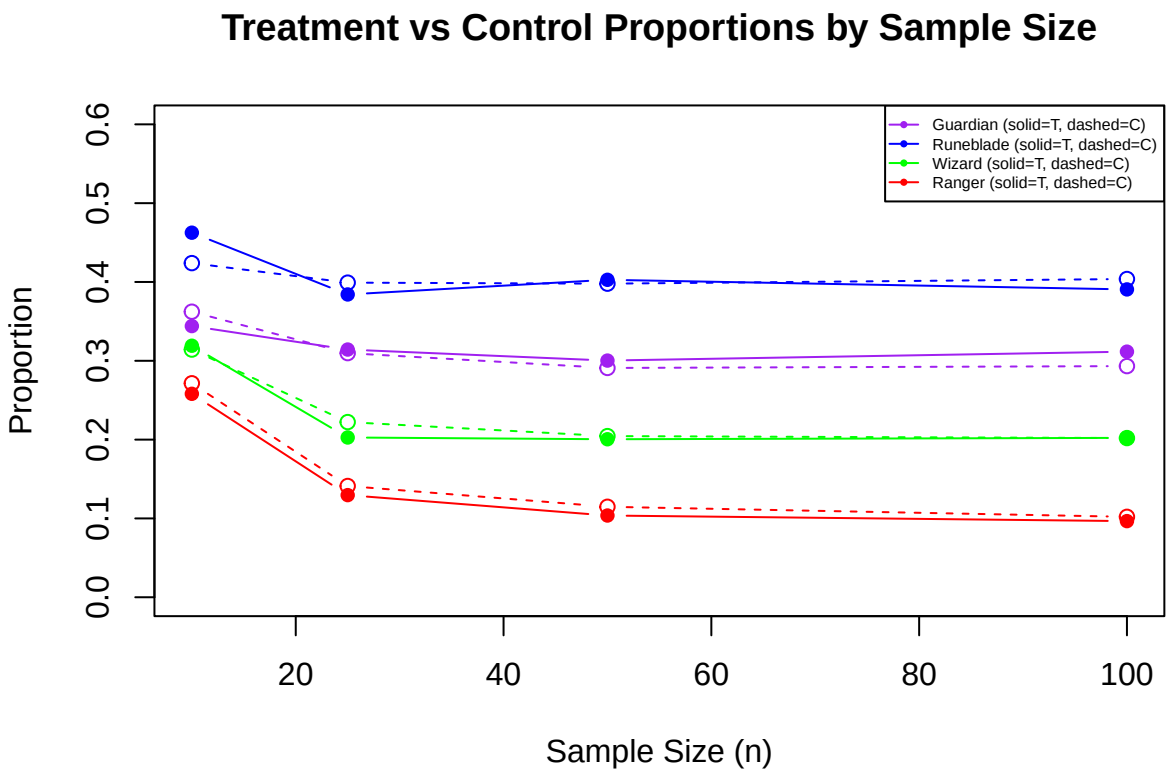
As we Increase the sample we converge onto the true values of .3, .4, .2, .1

**As n increases the treatment and control group approach each other**

```
#I like the graph visual here
plot(1, type='n', xlim=range(n_values),ylim=c(0,.6),
     xlab='Sample Size (n)', ylab='Proportion',
     main='Treatment vs Control Proportions by Sample Size')
roles<-unique(avg_results$Var1)
colors<-c('purple','blue','green','red')

for (i in seq_along(roles)) {
  treat_data<- avg_results[avg_results$Var1==roles[i] &
                             avg_results$GroupType=='Treatment',]
  control_data<- avg_results[avg_results$Var1==roles[i] &
                              avg_results$GroupType=='Control',]
  #line types
  lines(treat_data$SampleSize, treat_data$Freq, type='b',col=colors[i], pch=16)
  lines(control_data$SampleSize, control_data$Freq, type='b',col=colors[i], lty=2)
}

legend('topright',legend=paste(roles,'(solid=T, dashed=C)'),
       col=colors,lty=1, pch=16,cex=0.5)
```




## Data Analysis

```
#Read In data
setwd('C:/Users/natha/OneDrive/Documents/GitHub/NBW Pols 602 Work/Problem Set 1')
df <- read.csv('data/voting.csv')
voting<- df
```

### Question 1: Observations

**1a. The Treatment variable is whether or not they received the message or not.**

**1b. It is a discrete Variable**

**1c.The data type is Character, it is the word "yes" or "no"**

### Question 2: Binary Variable Assignment

```
#New binary Variable
#use ifelse for vectors
#not If()
#else()
voting$treated<- ifelse(voting$message == "yes",1,0)
```

## Question 3: Average Group Outcomes

```
#Need to subset the groups to get the means
Treated<-voting[voting$treated==1, ]
Control<-voting[voting$treated==0, ]
#now can look at group outcome
mean(Treated$voted)
```

```
## [1] 0.3779482
```

```
mean(Control$voted)
```

```
## [1] 0.2966383
```

```
#quick table
aggregate(voted ~ treated, data =  voting, mean)
```

```
##   treated     voted
## 1       0 0.2966383
## 2       1 0.3779482
```

```
#difference
treated_mean <- mean(Treated$voted)
control_mean <- mean(Control$voted)
Effect<- treated_mean - control_mean
Effect
```

```
## [1] 0.08130991
```

The likelihood that somebody voted, on average, if they received the experimental treatment (a mailed message) was 37.8 percent.

For the control group (those who did not receive a message) the likelihood the voted on average was 29.7 percent.

## Question 4: Creating new Data Frames from Subsets

```
#New treatment df
TreatmentGroup_df<- voting[voting$treated == 1,]
#new control df
ControlGroup_df<- voting[voting$treated== 0, ]
#check that no data dropped
nrow(voting)
```

```
## [1] 229444
```

```
nrow(TreatmentGroup_df)
```

```
## [1] 38201
```

```
nrow(ControlGroup_df)
```

```
## [1] 191243
```

```
nrow(ControlGroup_df) + nrow(TreatmentGroup_df)
```

```
## [1] 229444
```

## Question 5: Average Birth Years

```
mean(TreatmentGroup_df$birth)
```

```
## [1] 1956.147
```

```
mean(ControlGroup_df$birth)
```

```
## [1] 1956.186
```

Average Birth year for the treated group is 1956.147

Average Birth Year for the control group is 1956.186

## Question 6: Estimated Average Causal Effect

```
#Did this code already for question 3 so just display results
Effect
```

```
## [1] 0.08130991
```

There is a roughly 8 percentage point increase in the likelihood that an individual will vote if they received the message (treatment) compared to those who did not (control).

This would signal that social pressure does contribute to increased voting likelihood, at least under the design of this papers experiment.

## Question 7: Generalizing these results

In order for us to be able to generalize these results to the entire U.S population we would need to be able to assume that the survey population is not systematically different than the U.S. population.

Some concerns then are that the experiment surveys only households, completely ignoring the large number of individuals who live in apartments or other dwellings. Other factors such as whether the survey was conducted only in one or few states, and whether those state populations generalize.

Based solely off the knowledge that this was a survey of households we can say this systematically discriminates against apartment dwellers and other home types. Which raises concerns about generalization.