Abigail Ward\_ HW1

Abigail Ward

6/23/2021

### Part 1

### Question 1

Proportion of paralytic polio cases in the “Placebo” group and separately in the “NotInoculated” group in the “RandomizedControl” trial.

para\_placebo<-dat$Paralytic[2]/dat$Population[2]  
notInoc\_RandomControl<-dat$Paralytic[3]/dat$Population[3]   
para\_placebo

## [1] 0.0005714882

notInoc\_RandomControl

## [1] 0.000357166

### Question 2

Calculation of the probability of a draw that is greater than or equal to the observed value without simulation.

prob\_sucess<-(dat$Paralytic[2]+dat$Paralytic[3])/(dat$Population[2]+dat$Population[3])   
trials<-dat$Population[2]  
1-pbinom(dat$Paralytic[2]-1, trials, prob\_sucess)

## [1] 0.003240697

### Part 2

### Question 3

Proportions in each category for the RandomizedControl Vaccinated and the RandomizedControl Placebo groups.

vac\_paralytic<-dat.2$Vaccinated[1]/(sum(dat.2$Vaccinated))  
vac\_NotParalytic<-dat.2$Vaccinated[2]/(sum(dat.2$Vaccinated))  
vac\_False<-dat.2$Vaccinated[3]/(sum(dat.2$Vaccinated))  
  
placebo\_Paralytic<-dat.2$Placebo[1]/(sum(dat.2$Placebo))  
placebo\_NotParalytic<-dat.2$Placebo[2]/(sum(dat.2$Placebo))  
placebo\_False<-dat.2$Placebo[3]/(sum(dat.2$Placebo))  
  
vac\_paralytic

## [1] 0.402439

vac\_NotParalytic

## [1] 0.2926829

vac\_False

## [1] 0.304878

placebo\_Paralytic

## [1] 0.7098765

placebo\_NotParalytic

## [1] 0.1666667

placebo\_False

## [1] 0.1234568

### Question 4

Probability model for a simulation-based hypothesis test that addresses whether the two groups can reasonably be considered to come from populations with the same proportions of Paralytic, Non-Paralytic, and False Report.

#calculate the proportion of each result for both Vaccinated and placebo groups  
outcome.prop<-(dat.2$Vaccinated+dat.2$Placebo)/sum(dat.2$Vaccinated+dat.2$Placebo)  
outcome.prop

## [1] 0.6065574 0.2090164 0.1844262

#calculate the actual result  
prop.vacc<-dat.2$Vaccinated/sum(dat.2$Vaccinated)  
prop.vacc

## [1] 0.4024390 0.2926829 0.3048780

prop.plac<-dat.2$Placebo/sum(dat.2$Placebo)  
prop.plac

## [1] 0.7098765 0.1666667 0.1234568

test.stat<-sqrt(sum((prop.vacc-prop.plac)^2))  
test.stat

## [1] 0.3785652

#simulation  
n<-10000  
  
k1<-sum(dat.2$Vaccinated)  
k1

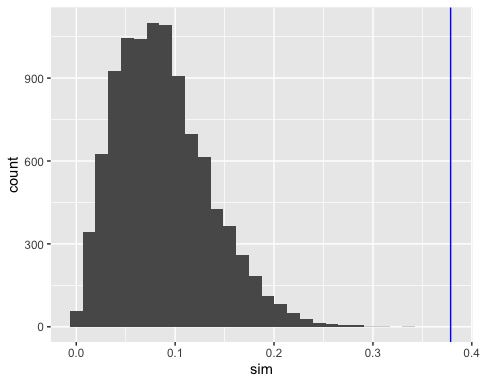
## [1] 82

k2<-sum(dat.2$Placebo)  
k2

## [1] 162

sim= rep(NA, n)  
for(i in 1:n){  
 samp1<-sample(c("Paralytic", "NonParalytic","FalseReports"), k1, replace=TRUE, prob = outcome.prop)  
 samp2<-sample(c("Paralytic", "NonParalytic","FalseReports"), k2, replace=TRUE, prob = outcome.prop)  
 counts1<-table(samp1)  
 counts2<-table(samp2)  
 props1<-counts1/sum(counts1)  
 props2<-counts2/sum(counts2)  
 dist.eu<-sqrt(sum((props1-props2)^2))  
   
 sim[i]<-dist.eu  
}  
dat.temp<-data.frame(sim=sim)  
ggplot(data =dat.temp, aes(x=sim))+geom\_histogram()+geom\_vline(xintercept=test.stat, color="blue")

## `stat\_bin()` using `bins = 30`. Pick better value with `binwidth`.



mean(sim>=test.stat)

## [1] 0

The distance between the proportion of Paralytic, Non-Paralytic and False Reports between the vaccinated and placebo groups is the test statistic. It is calculated by finding the euclidean distance between the vaccinated sample population and placebo sample population.

The probability model is the outcome.prop which describes the probability of each group (Paralytic, Non-Paralytic and False Reports) occurring in the population which determined using the observed values.

The probability model can be simulated by selecting an sample set of the three options (Paralytic, Non-Paralytic and False Reports) the same size and the placebo and the separately of the vaccinated groups based on the outcome.prop.

The comparison which best addresses the null hypothesis is done by calculating the number of times the simulation resulted in a test statistic equal to or greater than the observed statistic.