Polio Simulation, Shuffle Respondents

C. Durso

## Load Data

data("PolioTrials")  
dat<-PolioTrials  
#str(dat)  
kable(dat)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Experiment | Group | Population | Paralytic | NonParalytic | FalseReports |
| RandomizedControl | Vaccinated | 200745 | 33 | 24 | 25 |
| RandomizedControl | Placebo | 201229 | 115 | 27 | 20 |
| RandomizedControl | NotInoculated | 338778 | 121 | 36 | 25 |
| RandomizedControl | IncompleteVaccinations | 8484 | 1 | 1 | 0 |
| ObservedControl | Vaccinated | 221998 | 38 | 18 | 20 |
| ObservedControl | Controls | 725173 | 330 | 61 | 48 |
| ObservedControl | Grade2NotInoculated | 123605 | 43 | 11 | 12 |
| ObservedControl | IncompleteVaccinations | 9904 | 4 | 0 | 0 |

## Frame Question

The basic question is “did the vaccine work?” This can be answered in various ways. We would like to know whether the lower counts for the vaccinated population are consistent with chance, or whether there’s evidence of a structural difference.

Are the counts of paralytic polio cases consistent with the model that there were going to be 148 cases and these were randomly assigned to the “Vaccinated” and “Placebo” groups?

## Test

We can test this by drawing a random sample of the size of the vaccinated group from all the respondents then counting the number of paralytic polio cases in the sample.

### Simulation

Create a synthetic population of the size of the population in the experiment. Model respondents who didn’t contract paralytic polio with a “0”, those who did with “1”.

pop<-rep(0, sum(dat$Population[1:2]))  
ct<-sum(dat$Paralytic[1:2])  
pop[1:ct]<-1

Sample a possible vaccinated group. The sum of the values sampled will be the count of paralytic polio cases in the simulated “Vaccinated” population.

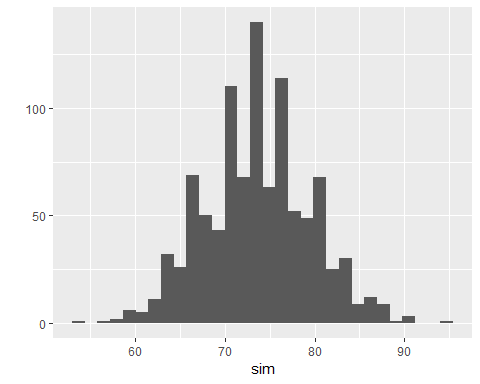
tot.vac<-dat$Population[1]  
sim<-sample(pop,tot.vac)  
sum(sim)

## [1] 68

One sample isn’t conclusive. Let’s loop through the sampling process many times. (Looping is often avoidable in R. See version below.)

n<-1000  
sim<-numeric(1000)  
set.seed(345689876)  
for( i in 1:n){  
 sim[i]<-sum(sample(pop,tot.vac))  
}  
qplot(sim)

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



mean(sim<dat$Paralytic[1])

## [1] 0

This is a vectorized version.

ct.sim<-function(){  
 return(sum(sample(pop,tot.vac)))  
}  
set.seed(345689876)  
sim2<-replicate(n,ct.sim())  
qplot(sim2)  
mean(sim2<dat$Paralytic[1])

Or we can rethink the model. For example we could take each case of paralytic polio and assign it to “Vaccinated” with probability equal to the proportion of vaccinated children in the study.