Inference Example, counts

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 “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?

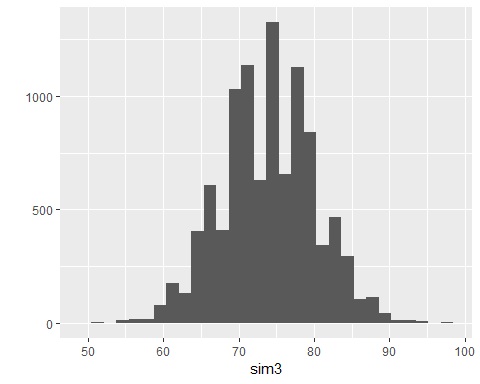
## Test

Take each case of paralytic polio and assign it to “Vaccinated” with probability equal to the proportion of vaccinated children in the study. Count the number assigned to “Vaccinated” and compare this with the observed number of polio cases in the “Vaccinated” group.

Here, rbinom(n,ct,prop) is a function that models the number of random assignments to the distinguished group from a population of size “ct” if the probability of assignment to the distinguished group is “prop”. The value of “n” is the number of times to repeat the experiment.

n<-10000  
ct<-sum(dat$Paralytic[1:2])  
prop<-dat$Population[1]/sum(dat$Population[1:2])  
set.seed(45678765)  
sim3<-rbinom(n,ct,prop)   
qplot(sim3)

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



mean(sim3<=dat$Paralytic[1])

## [1] 0

What does this say about the consistency of the observations with the chance model? (Is this the same as the first model?)