Problem Set 1 Solutions

D. Parada

January 16, 2019

Introduction

Questions are 10 points each.

These questions were rendered in R markdown through RStudio (https://www.rstudio.com/wp-content/uploads/2015/02/rmarkdown-cheatsheet.pdf, http://rmarkdown.rstudio.com).

Part 1

Question 1

Using the data from the polio trials and the simulation methods from inference_example_polio.Rmd, please calculate and display the proportion all polio cases to the study population for the "Vaccinated" and "Placebo" groups in the "RandomizedControl" experiment.

```
library(HistData)
```

```
## Warning: package 'HistData' was built under R version 3.4.4
```

```
dat<-PolioTrials
dat$tot<-dat$Paralytic+dat$NonParalytic
dat$tot[1:2]/dat$Population[1:2]
```

```
## [1] 0.0002839423 0.0007056637
```

Question 2

Using the probability model that the total number of all polio cases in the "Vaccinated" and "Placebo" group together is fixed and these were randomly assigned to the "Vaccinated" and "Placebo" groups according to the proportions of the study population in each group, please estimate the probability of obtaining a count of all polio cases for the "Vaccinated" group in the "RandomizedControl" experiment that is less than or equal to the observed number.

Please use the simulation methods from inference_example_polio.Rmd.

Please discuss whether the observed results are likely under this model, and relate this to the question of whether the vaccination was effective in reducing the total number of polio cases in a group.

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

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

```
# The results are not likely under this model. The total number of cases observed using this probabilit
```

Question 3

Consider the probability model that the number of all polio cases in the "Vaccinated" group of the "ObservedControl" experiment is a draw from the binomial distribution with the number of trials equal to the number of children in the "Vaccinated" group and the probability of "success" is equal to the proportion of polio cases in the "Vaccinated" and "Grade2NotInoculated" groups combined. Without simulation, calculate the probability of a draw that is less than or equal to the observed value.

```
size.this<-dat$Population[5]
prob.this<-(dat$tot[5]+dat$tot[7])/(dat$Population[5]+dat$Population[7])
pbinom(dat$tot[5],size=size.this,prob=prob.this)
```

```
## [1] 0.04221218
```

Question 4

Specify a null hypothesis with its probability model to address the question of whether the counts of total polio cases in the "Vaccinated" groups in both experiment are consistent with equal risk of polio in each group. Explain how to conduct and interpret the analysis based on the null hypothesis.

```
dat$tot[c(1,5)]/dat$Population[c(1,5)]
```

```
## [1] 0.0002839423 0.0002522545
```

One option is to estimate the risk by the proportion of the combined groups that contracted any form of polio. Then use this as the probability of contracting polio in a binomial model for the count of polio cases in each group. Note that the "Vaccinated" group from the "RandomizedControl" experiment has a higher rate than the "Vaccinated" group from the "ObservedControl" experiment. Calculate the probability of a count as high or higher in the "RandomizedControl" group. If this probability is very small, the observed results are unlikely under the null hypothesis. Because we don't have a structural reason to believe that the "RandomizedControl" group is at higher risk, the probability should be doubled for a two-tailed test.

(There are many other options.)

Question 5

Please implement the test you described in Question 4 and interpret the results.

```
size.this<-dat$Population[1]
prob.this<-(dat$tot[1]+dat$tot[5])/(dat$Population[1]+dat$Population[5])
(1-pbinom(dat$tot[1]-1,size=size.this,prob=prob.this)) # probability of dat$tot[1] or more
```

```
## [1] 0.3419806
```

The probability is large, so the observations are reasonably likely under the null hypothesis