

Assignment 2

Anant Vishwakarma

4/18/2022

The task was to estimate the sensitivity of a test for COVID-19 that uses as its reference sub sequence, the sub sequence from nucleotide 27894 to 28259 of the genome. Sensitivity is the number which represents true positives. Test sequences (reads) of length 3000 were used for each test, along with 12 random reads. 1000 replications of the test were carried out to ensure a better subset of results. After taking our reference sub sequence and comparing it to a random sequence of nucleotides from our data set we get the following results.

```
## repl1000
##      -   +
## 447 553
```

Discussion of Results

From this table, we would estimate that this test has a sensitivity of about $\frac{553}{447+553} \approx 0.55$.

This is the estimated probability with which we can get a positive result. However it should be known that these tests have been done under laboratory conditions and cannot account for all the uncertainties that might arise in real world settings but they still provide a useful estimate to further our understanding of the problem statement.

Appendix

```
knitr::opts_chunk$set(echo = FALSE)
knitr::opts_chunk$set(echo = FALSE)
CheckOnce <- function(FullSequence,TestRef,readLength){

  RandomIndex <- sample(1:(nchar(FullSequence) - readLength + 1),1)

  RandomSnippet <- substr(FullSequence, RandomIndex, RandomIndex + readLength - 1)

  grepl(TestRef,RandomSnippet)
}

Test <- function(FullSequence,TestRef,readLength,numReps){

  Repls <- replicate(numReps,CheckOnce(FullSequence,TestRef,readLength))
  ifelse(any(Repls),"+","-")
}
knitr::opts_chunk$set(echo = FALSE)
SeqLines <- readLines("http://www.stat.ucdavis.edu/~affarris/CovidRef.txt")
CovSequence <- paste(SeqLines[-1], collapse="")

RefSubseq <- substr(CovSequence,27894,28259)

repl1000 <- replicate(1000,Test(CovSequence,RefSubseq,3000,12))
results <- table(repl1000)
results
```