COVID-19-Clinical-Trial

Ali Amini

May 4, 2020

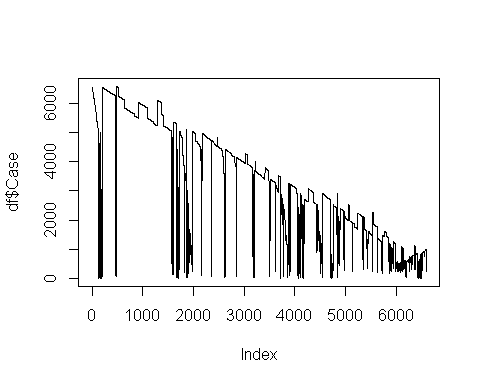
let’s load the data

df = data.frame(read.csv("COVID\_19\_SG.csv"))

glimpse(df)

## Observations: 6,588  
## Variables: 14  
## $ Case <int> 6560, 6558, 6531, 6529, 6509, 6497, 648...  
## $ Patient <fct> 37 Year Old Male Bangladeshi, 37 Year O...  
## $ Age <int> 37, 37, 31, 43, 31, 30, 23, 26, 36, 28,...  
## $ Gender <fct> Male, Male, Male, Male, Male, Male, Mal...  
## $ Nationality <fct> Bangladeshi, Bangladeshi, Bangladeshi, ...  
## $ Status <fct> In hospital, In hospital, In hospital, ...  
## $ Infection.Source <fct> Local transmission, Local transmission,...  
## $ Country.of.Origin <fct> Unclear origin, Unclear origin, Unclear...  
## $ Symptomatic.To.Confirmation <fct> -, -, -, -, -, -, -, -, -, -, -, -, -, ...  
## $ Days.To.Recover <fct> -, -, -, -, -, -, -, -, -, -, -, -, -, ...  
## $ Symptomatic.At <fct> -, -, -, -, -, -, -, -, -, -, -, -, -, ...  
## $ Confirmed.At <fct> "19th, Apr 2020", "19th, Apr 2020", "19...  
## $ Recovered.At <fct> -, -, -, -, -, -, -, -, -, -, -, -, -, ...  
## $ Displayed.Symptoms. <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRU...

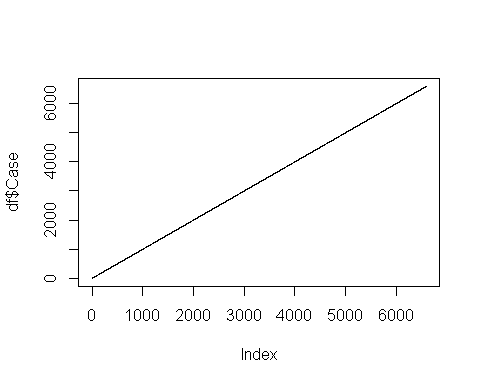
plot(df$Case, type = "l")



Let’s reorder the data frame based on the case number.

df <- df[order(df$Case),]

plot(df$Case, type = "l")



**Data cleaning:**

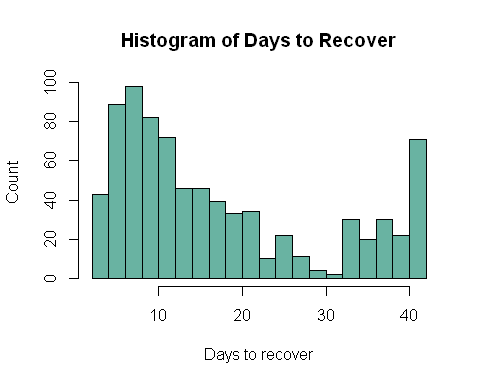
df[df == "-"]<- NA

glimpse(df)

## Observations: 6,588  
## Variables: 14  
## $ Case <int> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, ...  
## $ Patient <fct> 66 year old Wuhan Man, 53 year-old fema...  
## $ Age <int> 66, 53, 37, 36, 56, 56, 35, 56, 56, 56,...  
## $ Gender <fct> Male, Female, Male, Male, Female, Male,...  
## $ Nationality <fct> Chinese, Chinese, Chinese, Chinese, Chi...  
## $ Status <fct> Recovered, Recovered, Recovered, Recove...  
## $ Infection.Source <fct> Imported case, Imported case, Imported ...  
## $ Country.of.Origin <fct> China, China, China, China, China, Chin...  
## $ Symptomatic.To.Confirmation <fct> 3, 3, 1, 3, 4, 3, 4, 5, 6, 9, 3, 4, 3, ...  
## $ Days.To.Recover <fct> 26, 14, 27, 17, 21, 23, 7, 20, 25, 10, ...  
## $ Symptomatic.At <fct> "21st, Jan 2020", "21st, Jan 2020", "23...  
## $ Confirmed.At <fct> "23rd, Jan 2020", "24th, Jan 2020", "24...  
## $ Recovered.At <fct> "19th, Feb 2020", "7th, Feb 2020", "21s...  
## $ Displayed.Symptoms. <lgl> TRUE, TRUE, TRUE, TRUE, TRUE, TRUE, TRU...

df$Days.To.Recover<- as.numeric(df$Days.To.Recover)

hist(df$Days.To.Recover, breaks = 25, xlab = "Days to recover",  
 ylab = "Count", main = "Histogram of Days to Recover", col = "#69b3a2")



summary(df$Days.To.Recover)

## Min. 1st Qu. Median Mean 3rd Qu. Max. NA's   
## 2.00 8.00 13.00 17.82 25.00 42.00 5784

df %>%  
 filter(!is.na(Days.To.Recover))%>%  
 summarise(Standard\_deviaiton = sd(Days.To.Recover))

## Standard\_deviaiton  
## 1 12.28066

SD <- sd(df$Days.To.Recover, na.rm=TRUE)  
n <- 250

SE = (2\*SD^2/n)^0.5

X = seq(-16,16,0.1)  
Null = dnorm(X, sd = SE)  
Alt = dnorm(X, mean = -3, sd = SE)  
data <- data.frame(X, Null, Alt)

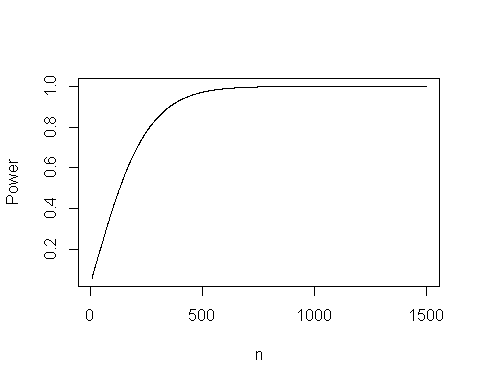
Z\_star\_95 = qnorm(0.025)\* SE

qnorm(0.8)-qnorm(0.025)

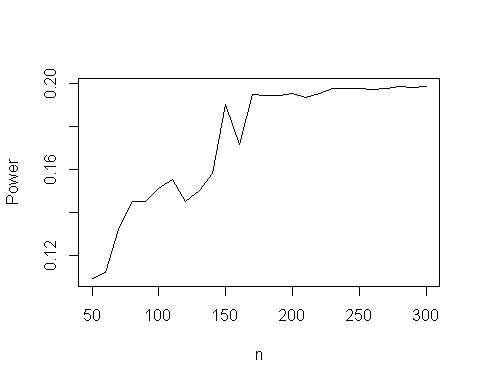
## [1] 2.801585

n <- seq(5,1500, by = 1)  
  
SE = (2\*SD^2/n)^0.5  
  
Z95 = qnorm(0.025)\* SE  
  
Power = pnorm(Z95, -3, SE)

plot(n,Power, type = "l")



n = seq(50,300,10)  
Power = c()  
j = 1  
for (i in n){  
simdf <- df %>%  
 filter(!is.na(Days.To.Recover))%>%  
 select(Case, Days.To.Recover)%>%  
 rep\_sample\_n(size = i, reps = 1000, replace = TRUE)%>%  
 summarise(Median = median(Days.To.Recover))  
B <- mean(simdf$Median) - quantile(simdf$Median, 0.025)  
A <- 3-B  
Power[j] <- sum(simdf$Median <= mean(simdf$Median)+A)/5000  
j <- j+1  
}  
plot(n, Power, type = "l")



#glimpse(simdf)  
#hist(simdf$Median, breaks = 60)