State_project

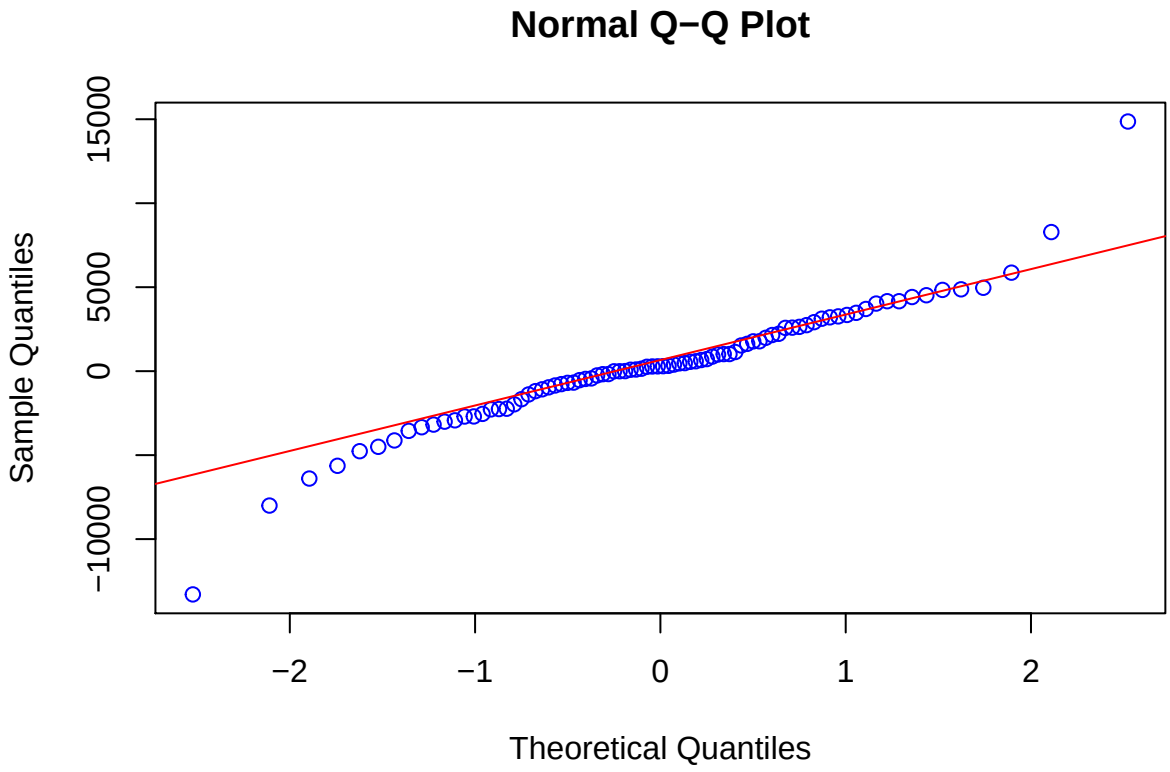
Kavya Gajjar

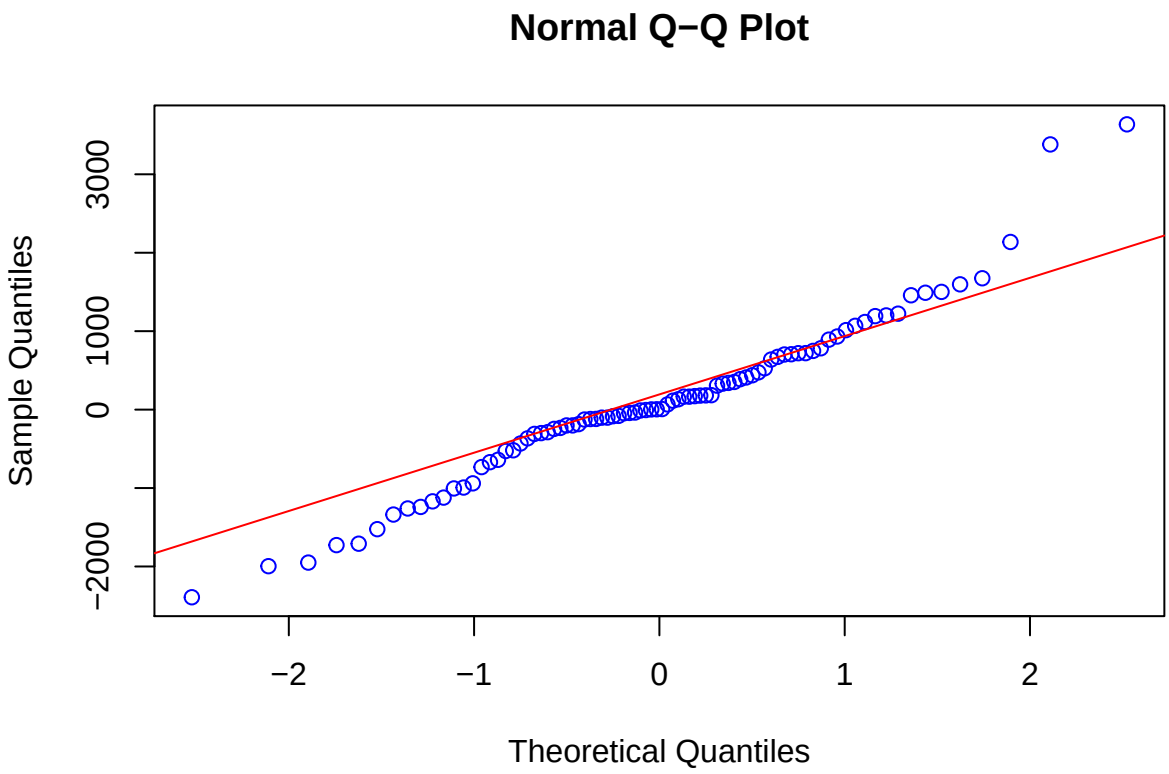
18/04/2021

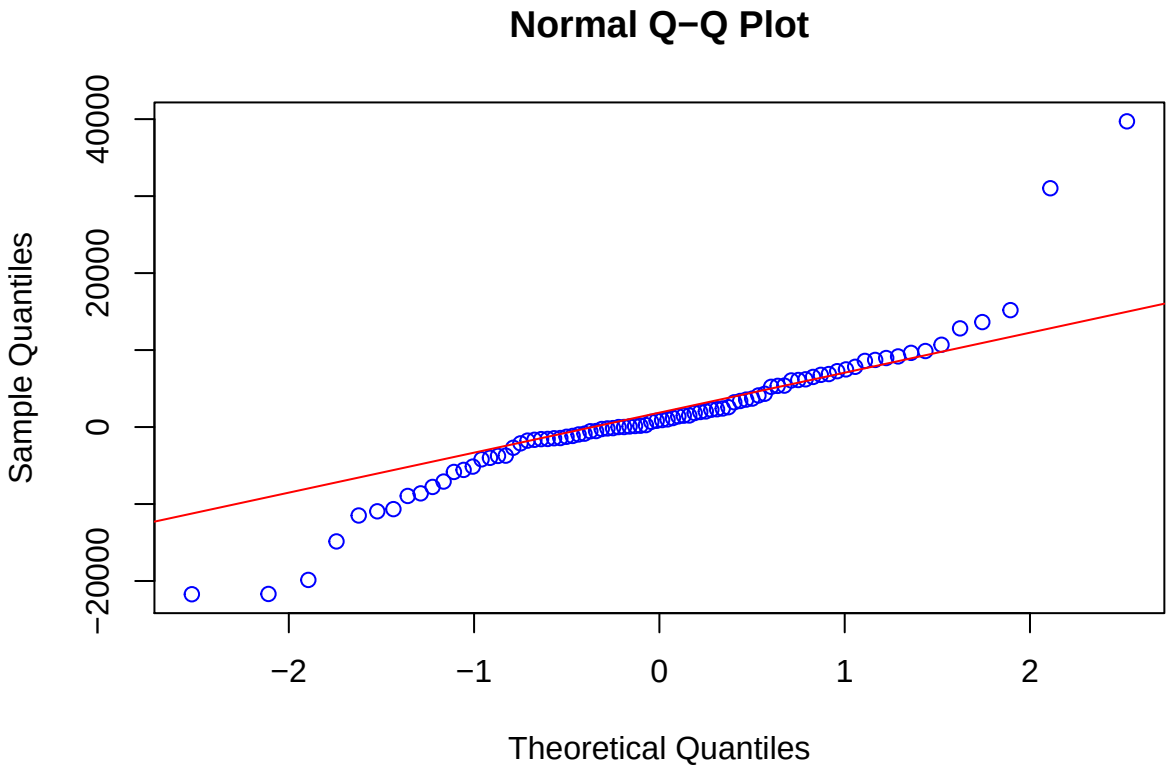
```
folder <- "D:\\MS DS\\DS 5110\\project\\Final project\\data\\"
state_data<- read_csv(paste(folder,"us_state_vaccinations.csv",sep=""))
globe_data <- read_csv(paste(folder,"country_vaccinations.csv",sep=""))
pop_data <- read_csv(paste(folder, "statewise_population.csv", sep="")) %>%
  select(State, Pop)
# source: https://worldpopulationreview.com/states


data <- state_data %>%
  filter(location != 'United States') %>%
  select(c('date','location','people_fully_vaccinated'))

data$date <- strptime(data$date, "%Y-%m-%d" )
data$date <- as.POSIXct(data$date)
data$people_fully_vaccinated <- as.numeric(data$people_fully_vaccinated)
colnames(pop_data)<- c("location", "population")

data <- inner_join(data, pop_data,by = c("location"))
data$population <- as.numeric(data$population)
```

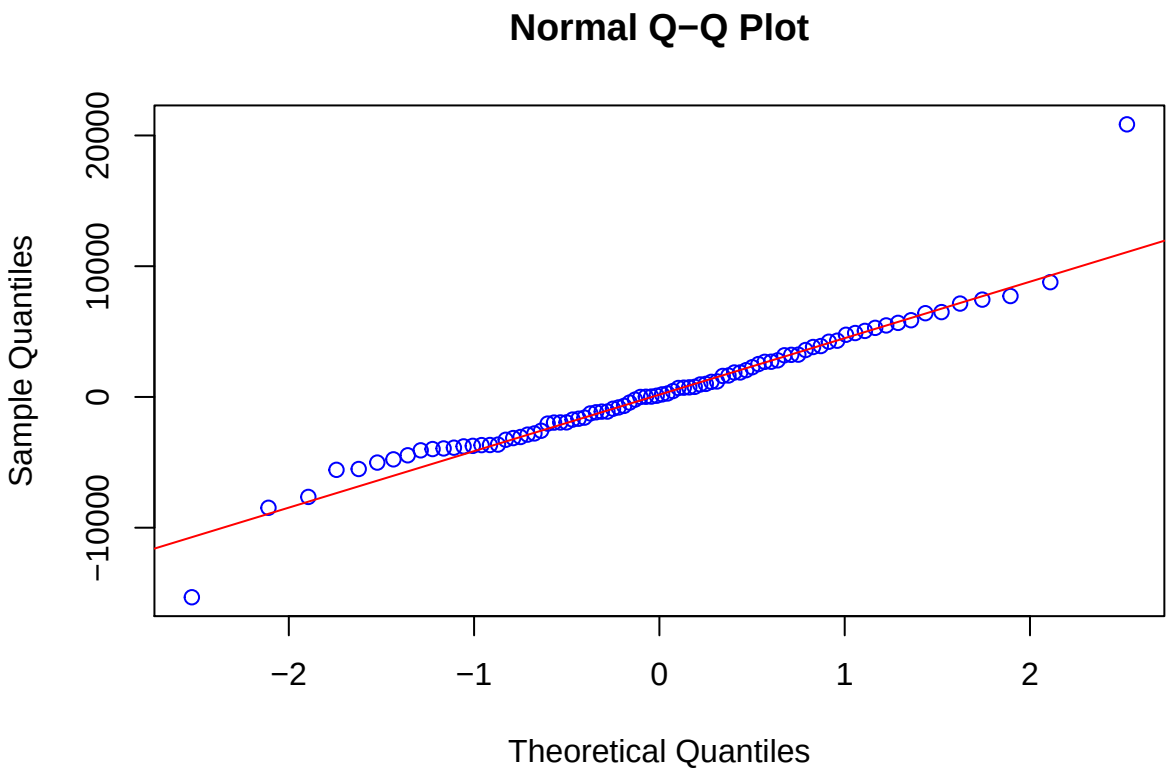
```
states <- unique(data$location)

output <- as.data.frame(states)
output$herd_immunity <- ''
for (name in states){
  #filter by state
  df <- data %>%
    filter(location == name)

  df$people_fully_vaccinated <- na.interpolation(df$people_fully_vaccinated,
                           option = "stine")
  training <- window(df$people_fully_vaccinated)
  arima_optimal = auto.arima(training)
    # arima(training, order = c(1,2,3))
    # auto.arima(training)
  # print(arima_optimal)
  my_fc <- forecast(arima_optimal, h = 400)
  pop <- df$population[1]
  herd <- pop * 0.9
  num = 0
  for(i in 1:length(my_fc$mean)){
```

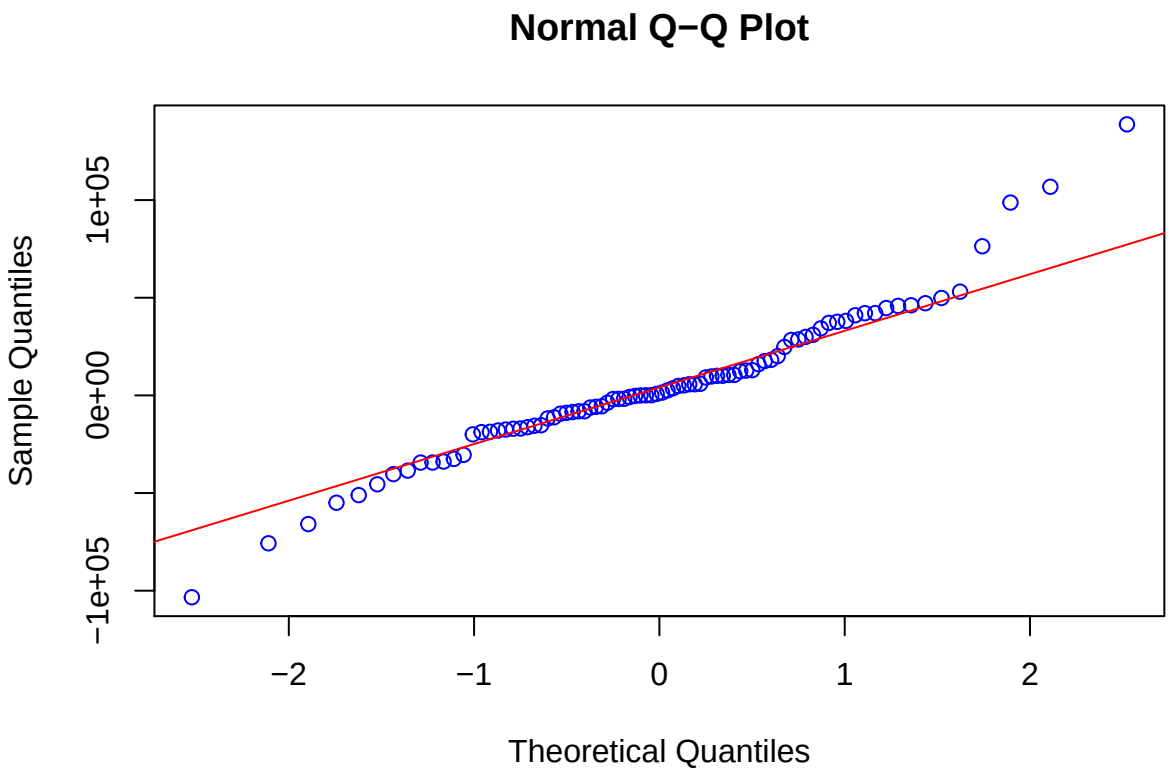
```
    if (my_fc$mean[i] >= herd){
      num = i
      break
    }
  }


  date <- as.Date(max(df$date))
  final <- date + num
  final = format(final, "%d,%B, %Y")
  # print(paste("Herd Immunity for ",name," can be achieved by ",
  #       final, sep=""))
  output$herd_immunity[output$states == name] <- final

  qqnorm(arima_optimal$residuals, col = 'blue')
  qqline(arima_optimal$residuals, col = 'red')
}
```

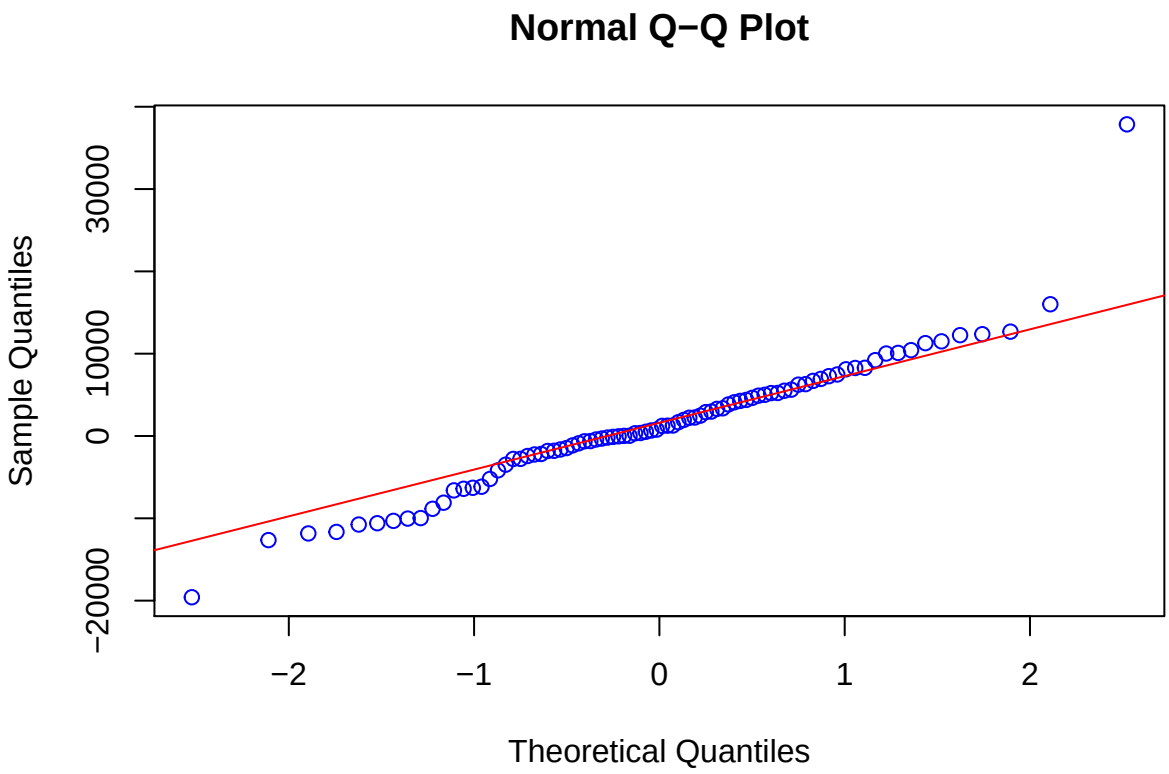


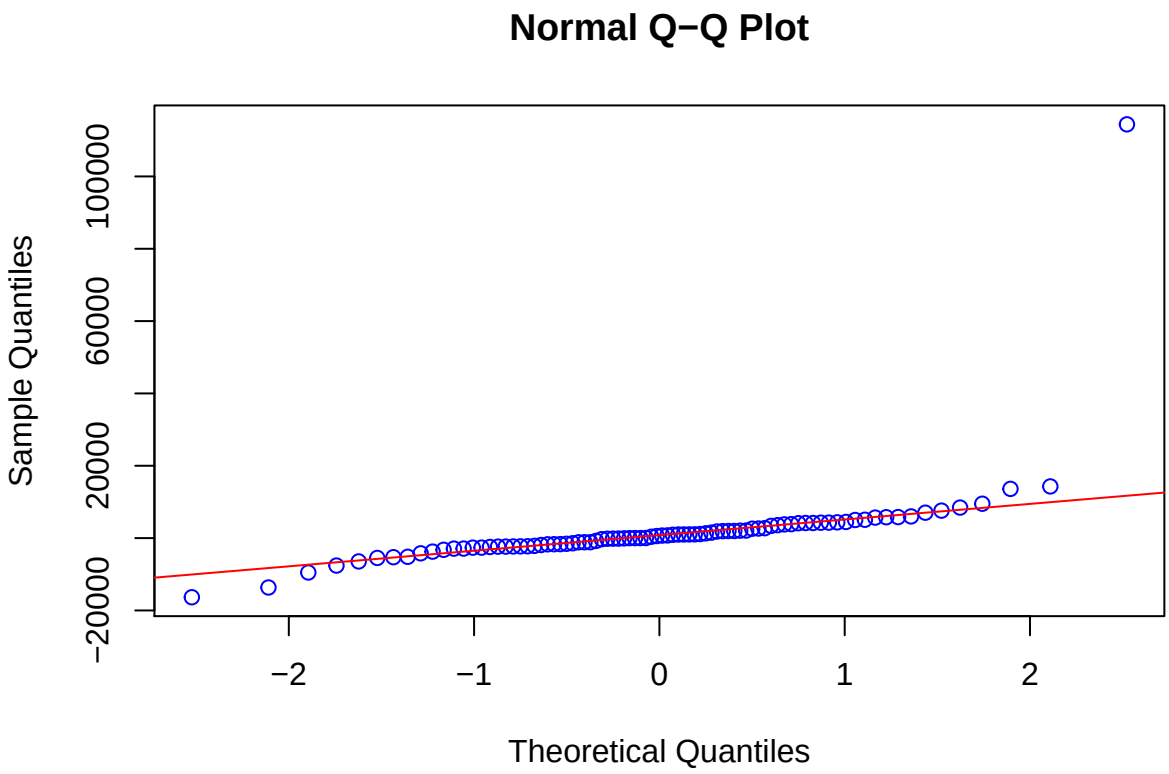

Normal Q−Q Plot
Sample Quantiles
3000
1000
0
−2000
−2
−1
0
1
2
Theoretical Quantiles

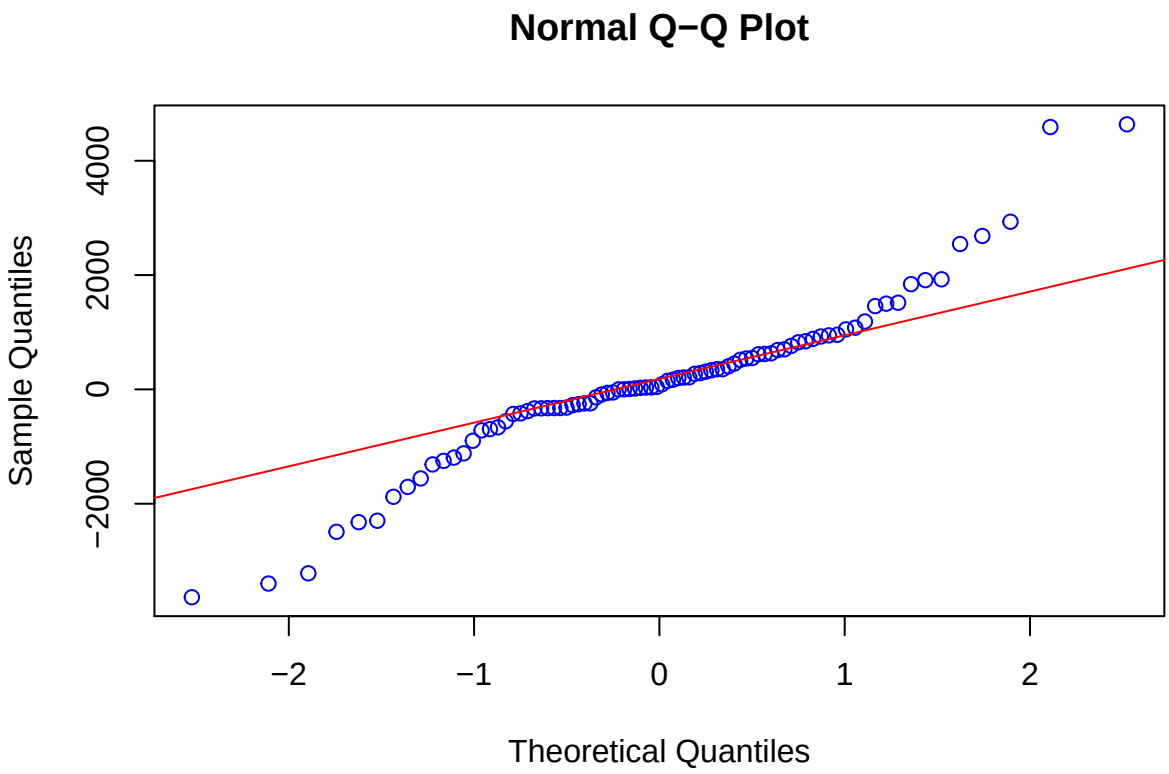

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
40000
20000
0
−20000
−2
−1
0
1
2

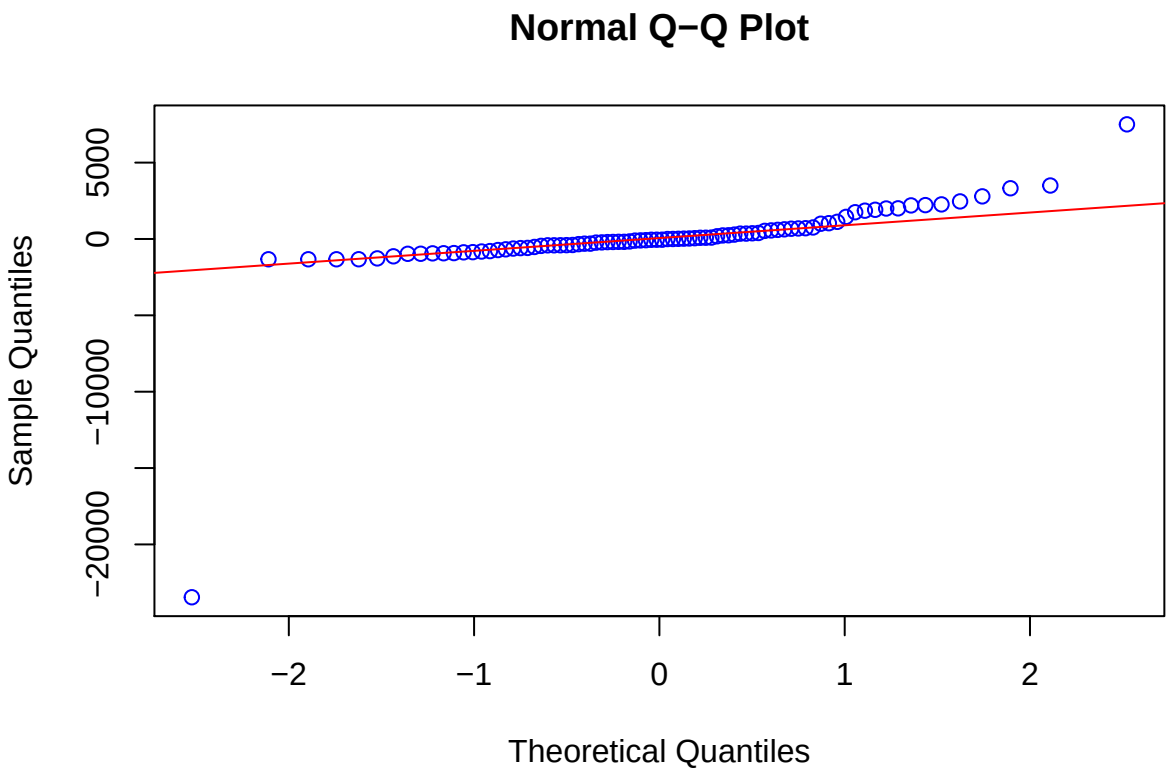

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
20000
10000
0
−10000
−2
−1
0
1
2

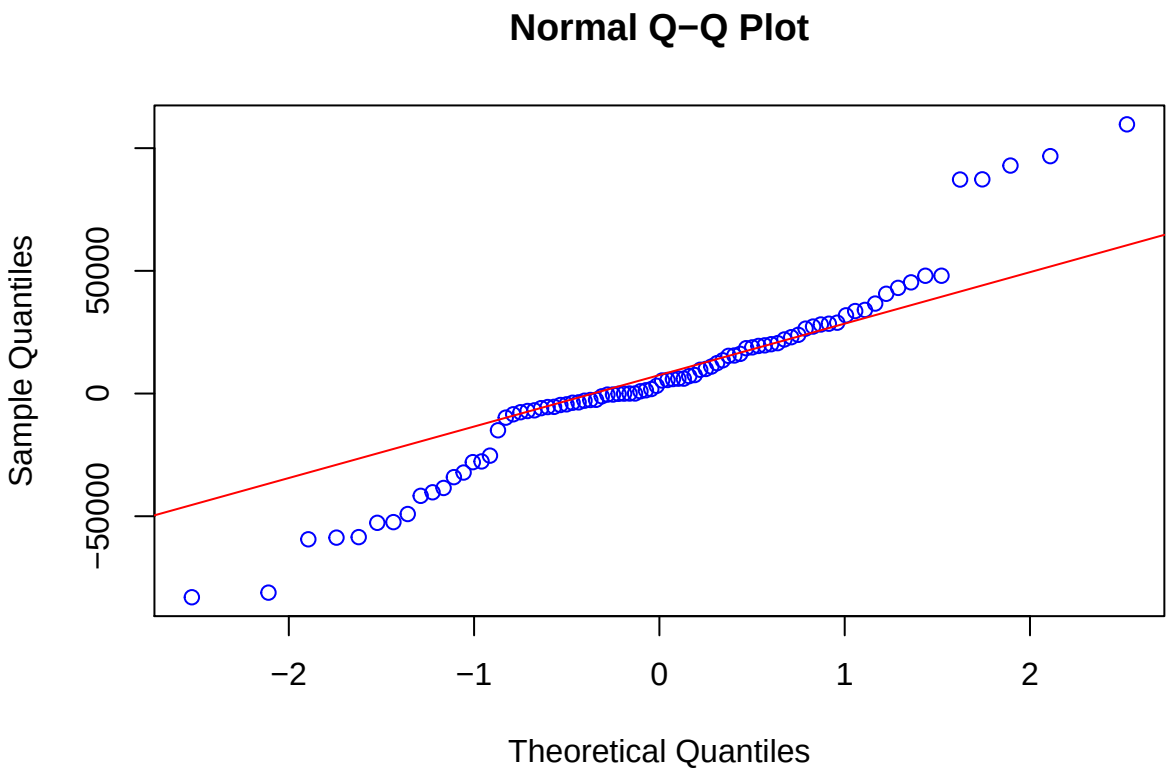

Normal Q−Q Plot
Sample Quantiles
1e+05
0e+00
−1e+05
−2
−1
0
1
2
Theoretical Quantiles

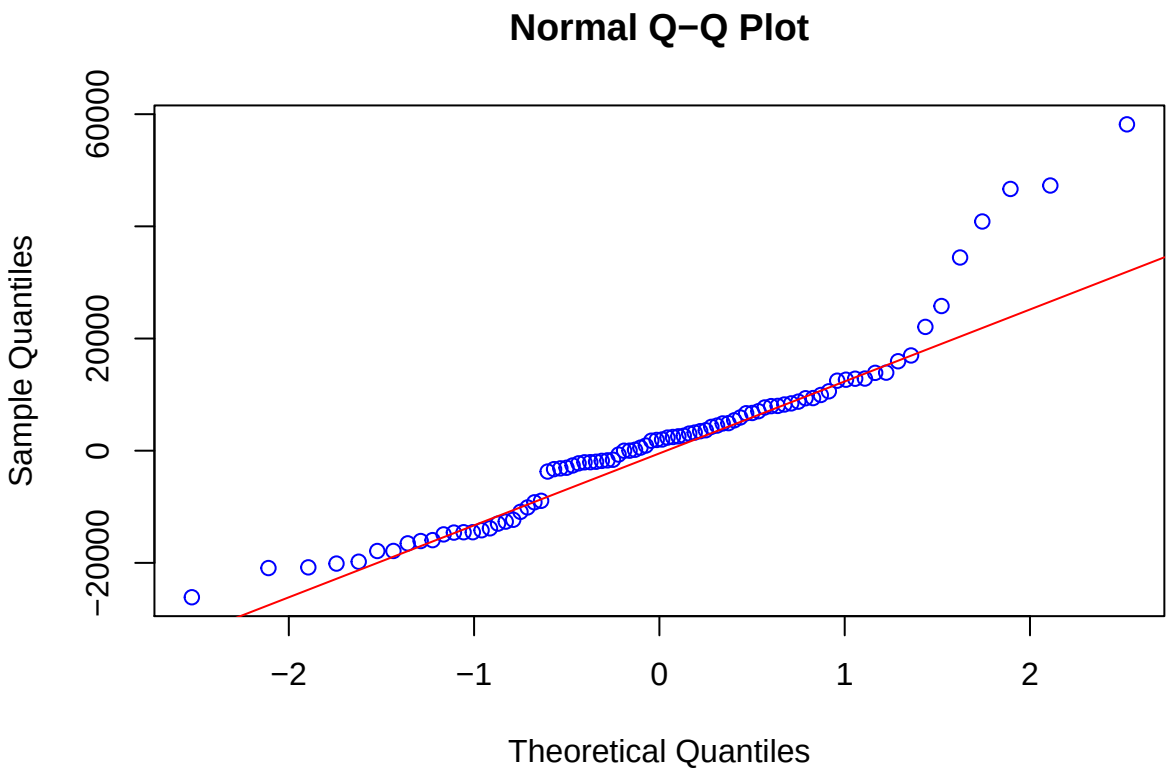

Normal Q−Q Plot
Sample Quantiles
30000
10000
0
−20000
−2
−1
0
1
2
Theoretical Quantiles

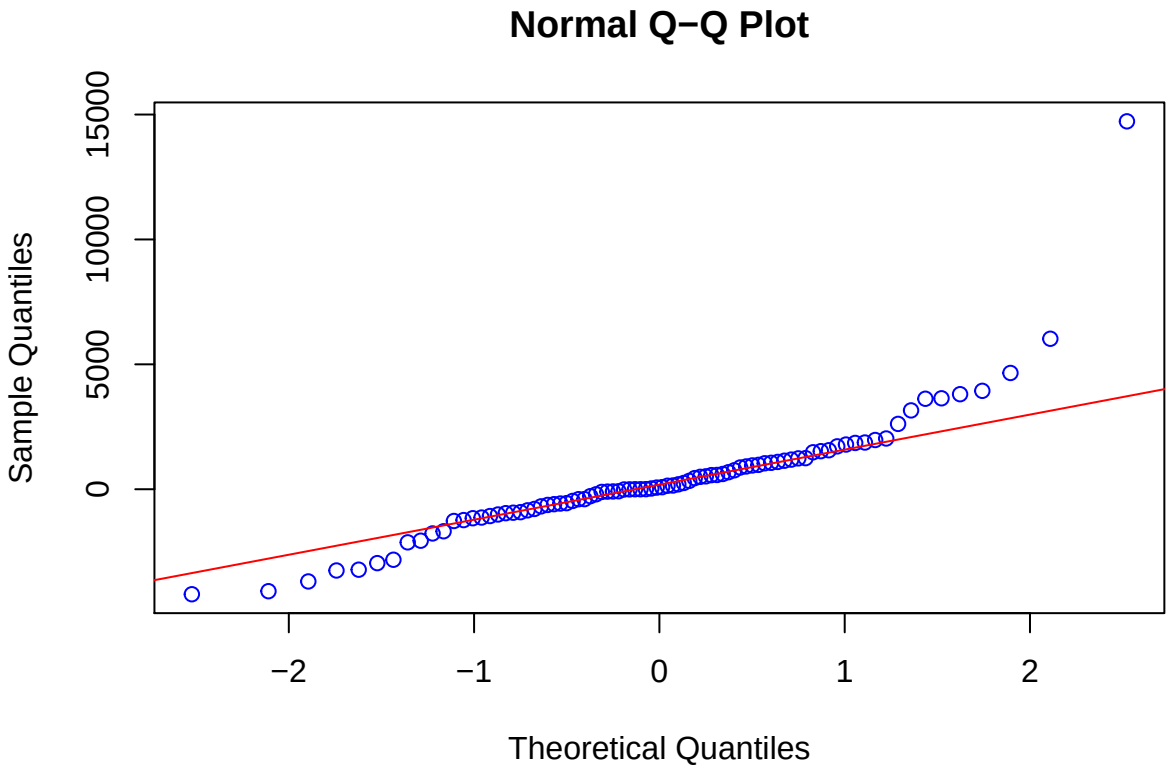

Normal Q−Q Plot
Sample Quantiles
−20000
20000
60000
100000
−2
−1
0
1
2
Theoretical Quantiles

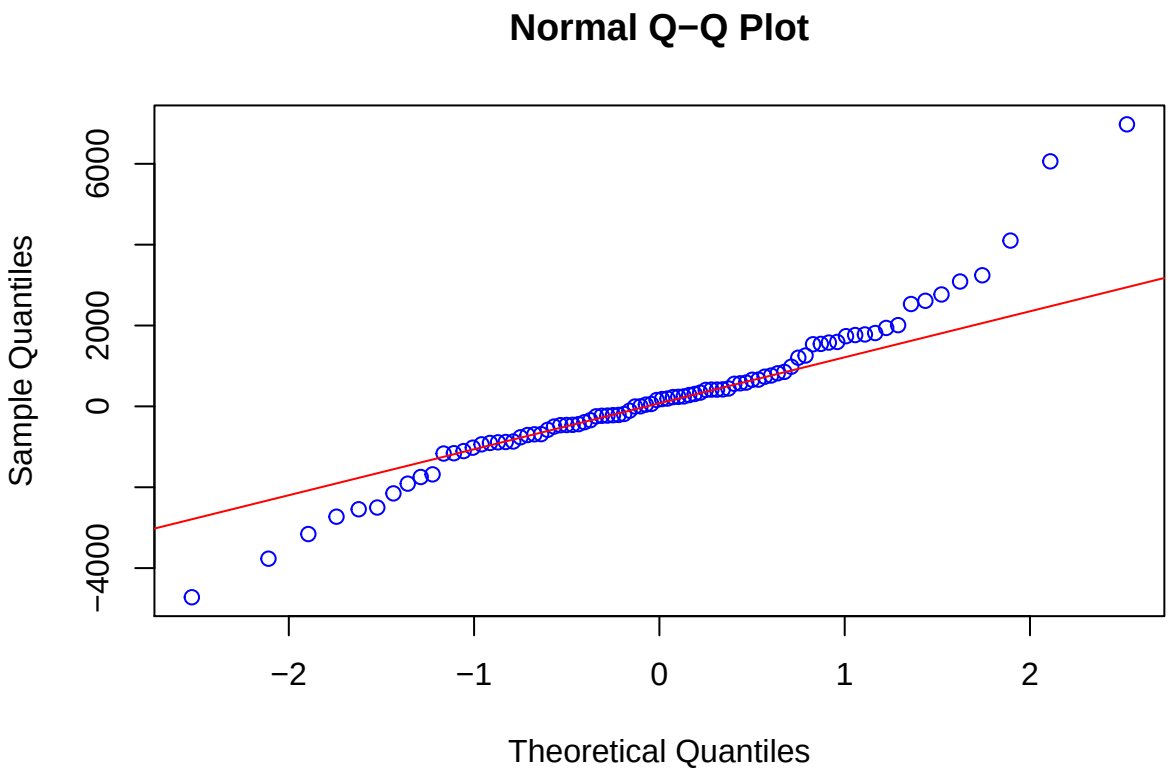

Normal Q−Q Plot
Sample Quantiles
−2000
0
2000
4000
−2
−1
0
1
2
Theoretical Quantiles

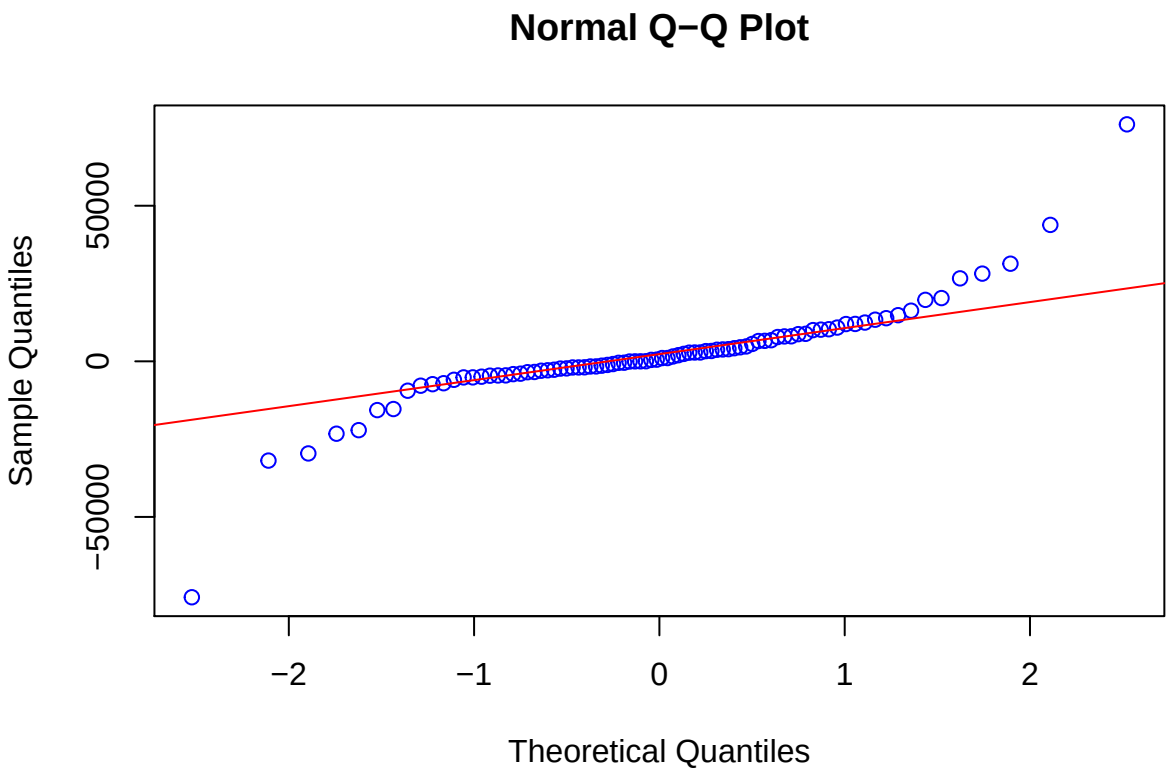

Normal Q−Q Plot
Sample Quantiles
5000
0
−10000
−20000
−2
−1
0
1
2
Theoretical Quantiles

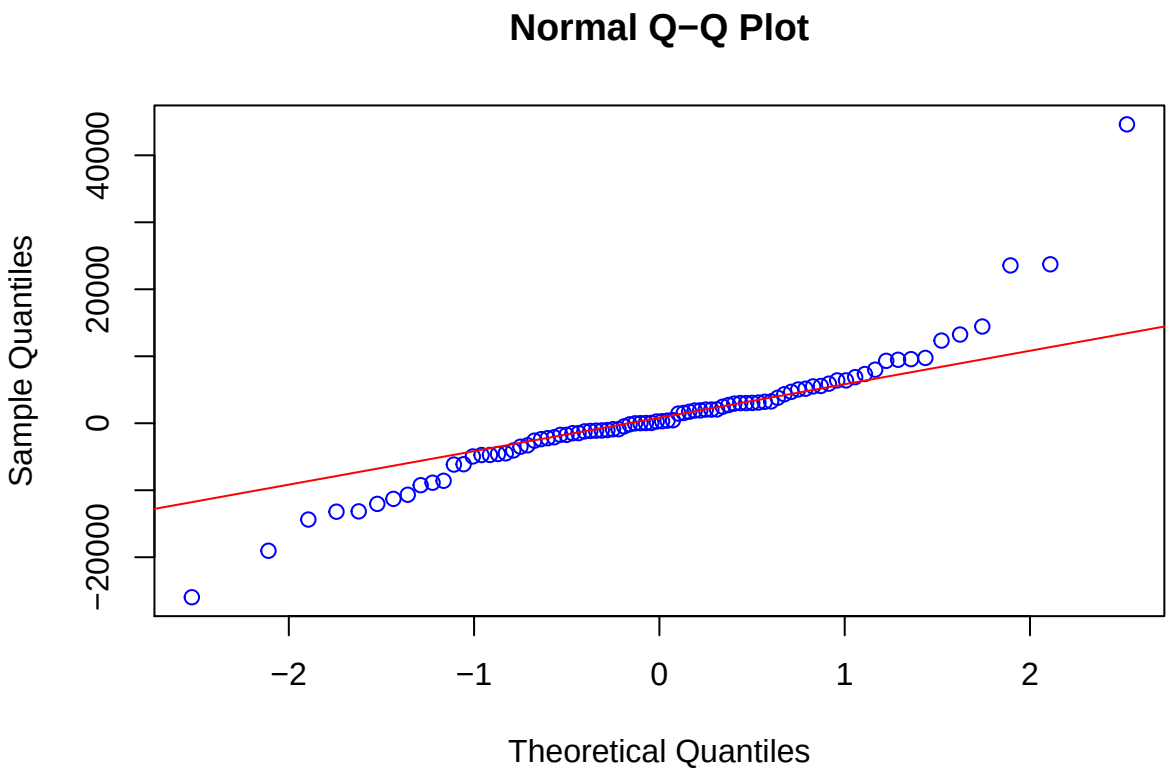

Normal Q−Q Plot
Sample Quantiles
−50000
0
50000
−2
−1
0
1
2
Theoretical Quantiles

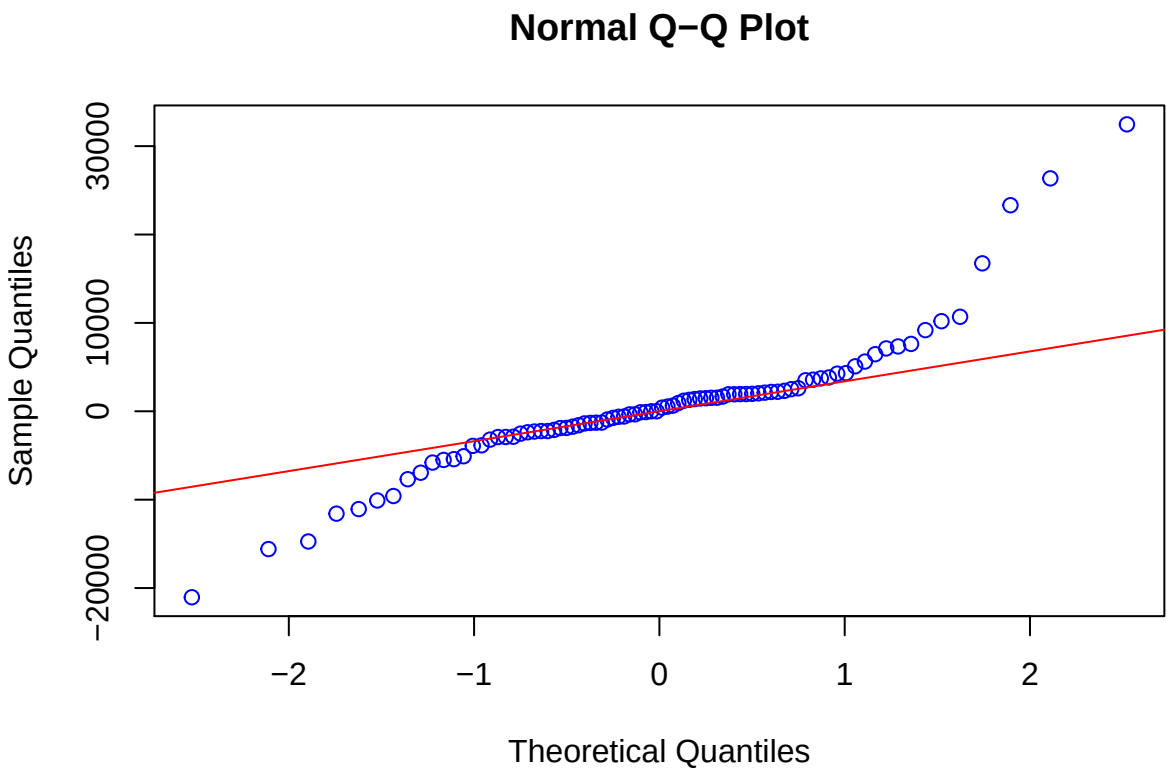

Normal Q−Q Plot
Sample Quantiles
60000
20000
0
−20000
−2
−1
0
1
2
Theoretical Quantiles

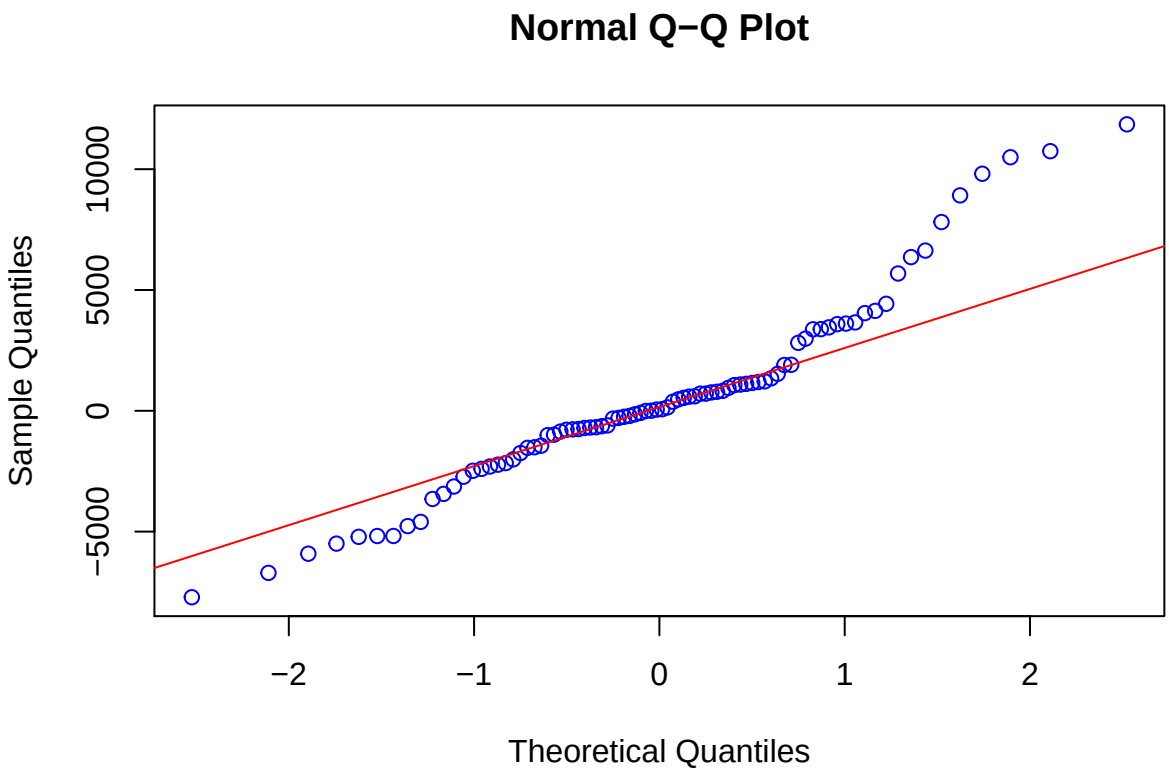

Normal Q−Q Plot
Sample Quantiles
0
5000
10000
15000
−2
−1
0
1
2
Theoretical Quantiles

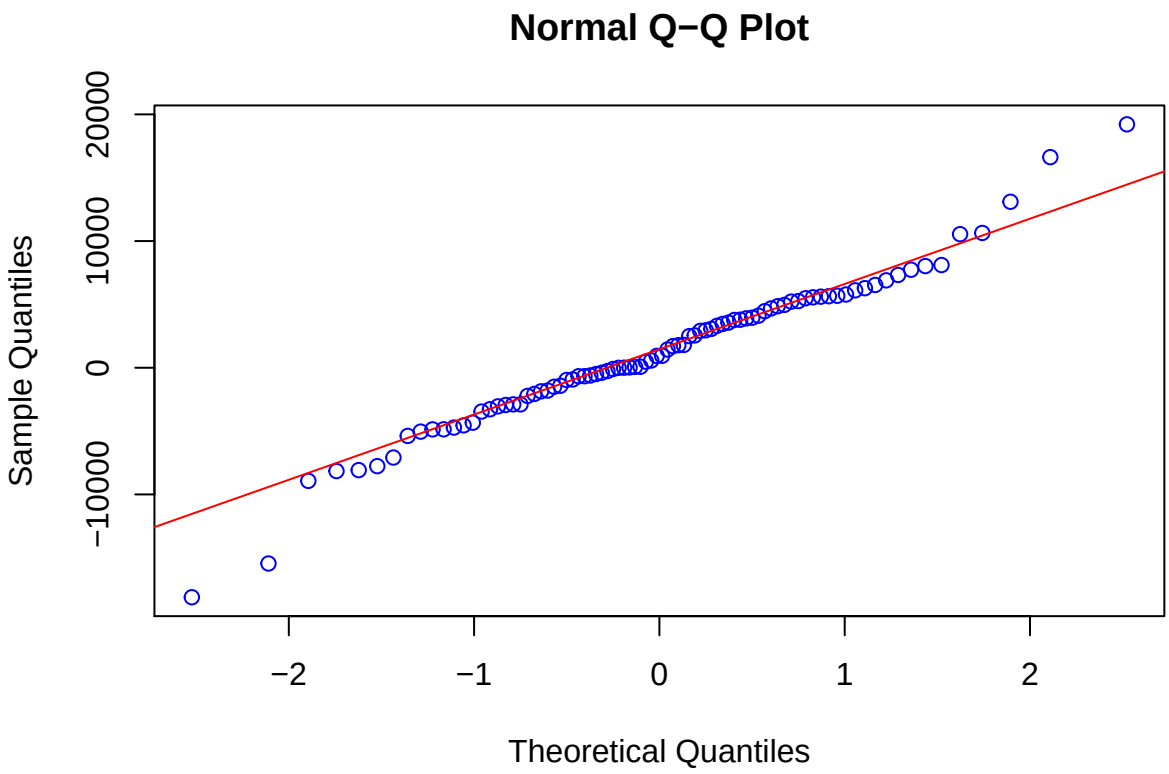

Normal Q−Q Plot
Sample Quantiles
6000
2000
0
−4000
−2
−1
0
1
2
Theoretical Quantiles

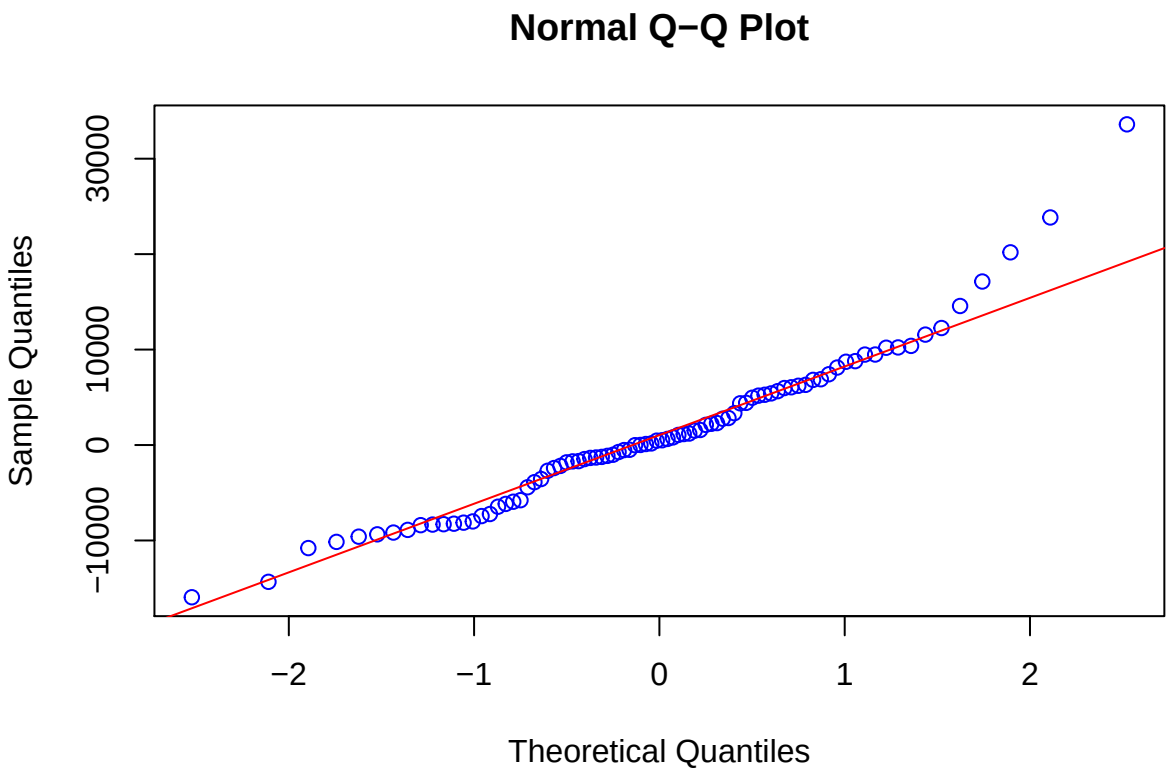

Normal Q−Q Plot
Sample Quantiles
50000
0
−50000
−2
−1
0
1
2
Theoretical Quantiles

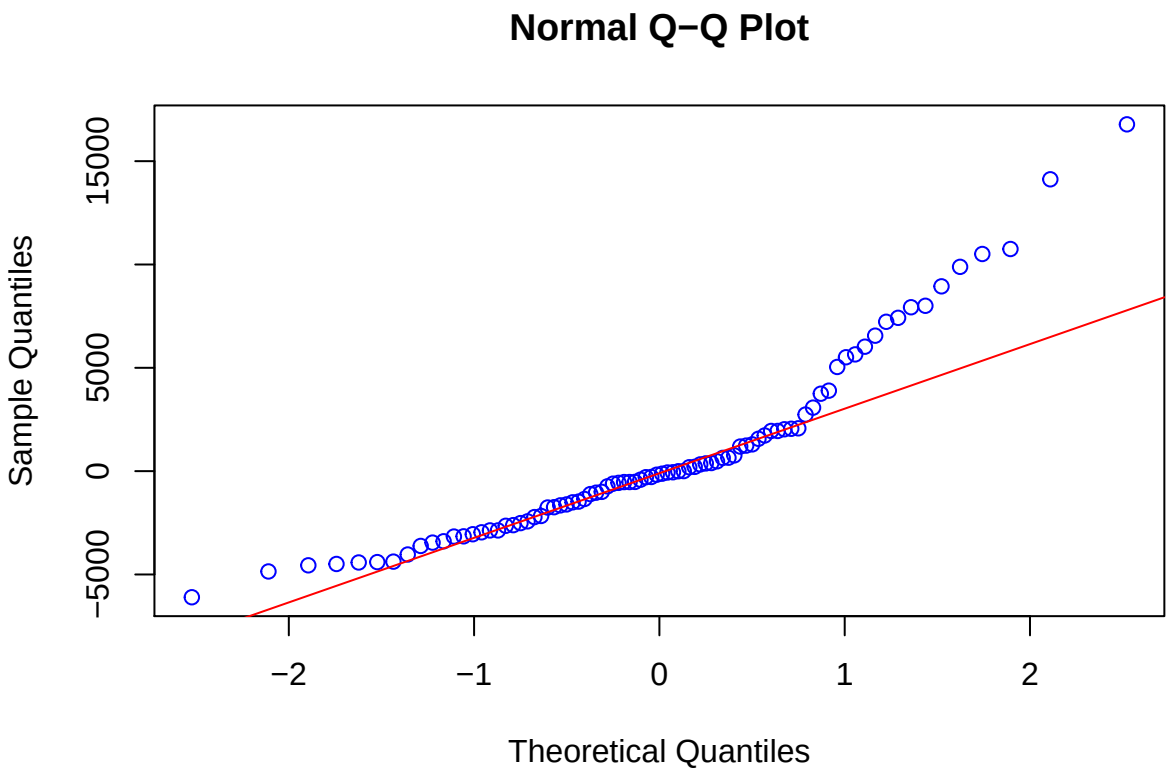

Normal Q−Q Plot
Sample Quantiles
40000
20000
0
−20000
−2
−1
0
1
2
Theoretical Quantiles

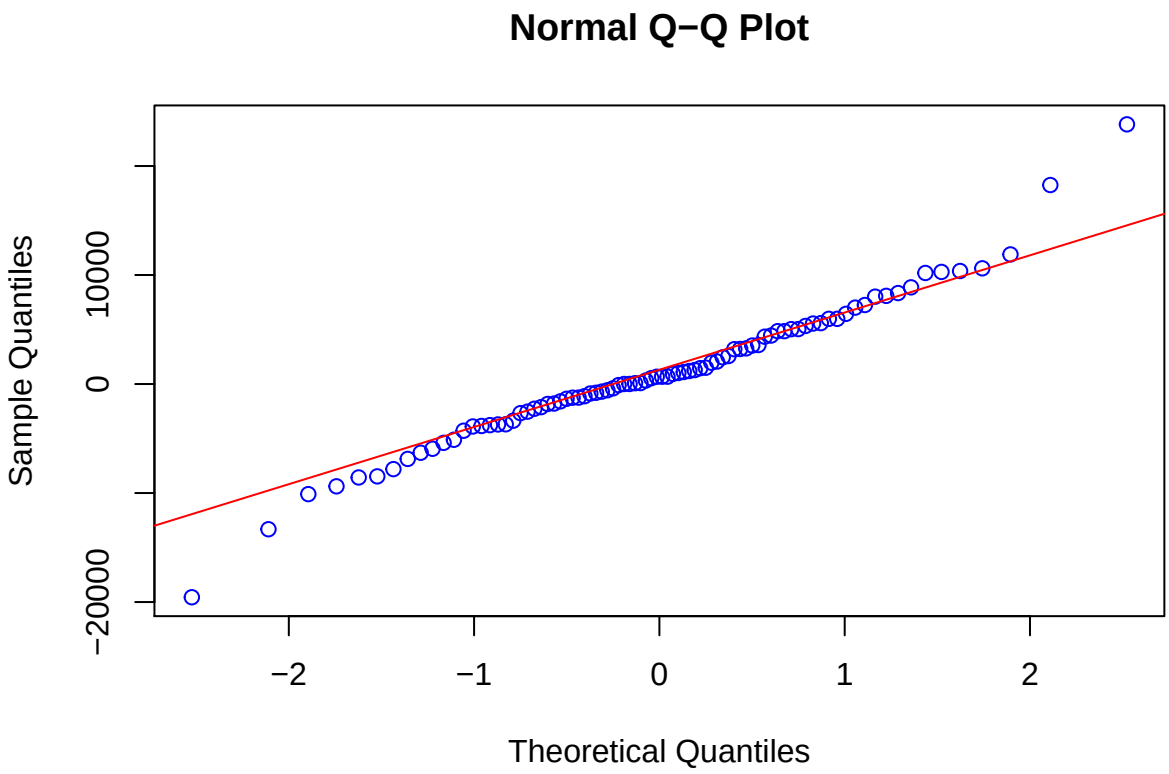

Normal Q−Q Plot
Sample Quantiles
30000
10000
0
−20000
−2
−1
0
1
2
Theoretical Quantiles

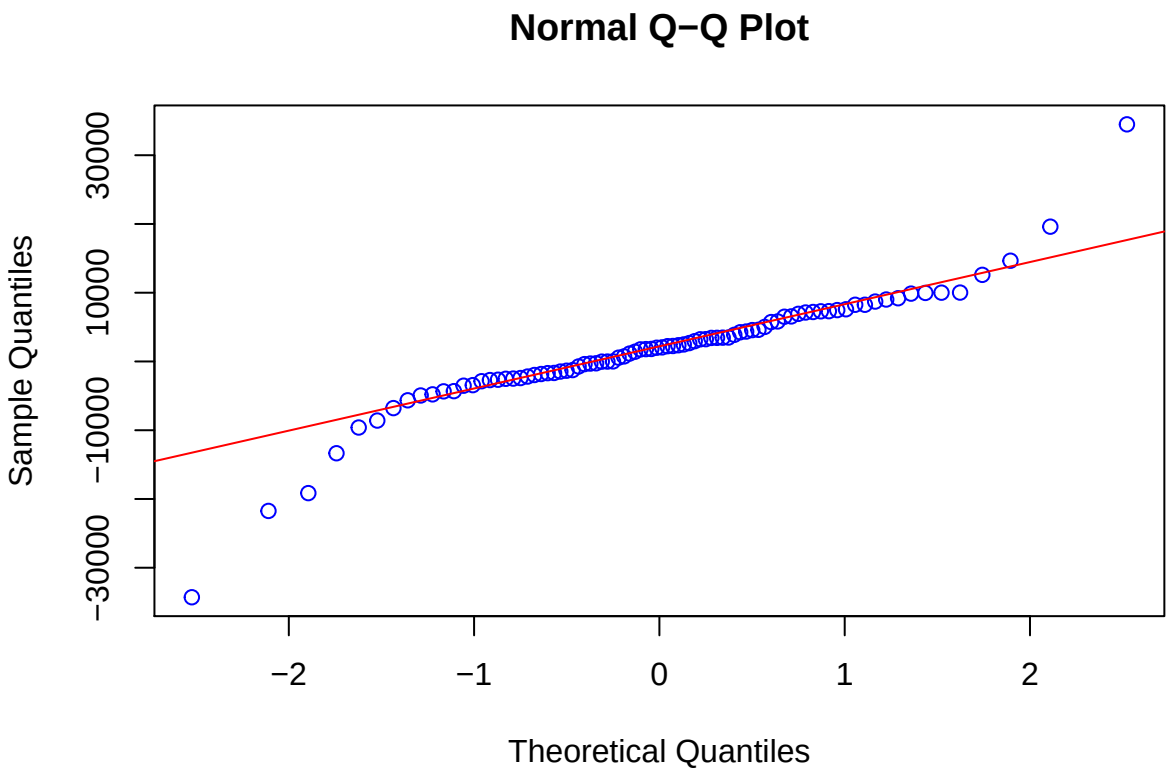

Normal Q−Q Plot
Sample Quantiles
−5000
0
5000
10000
−2
−1
0
1
2
Theoretical Quantiles

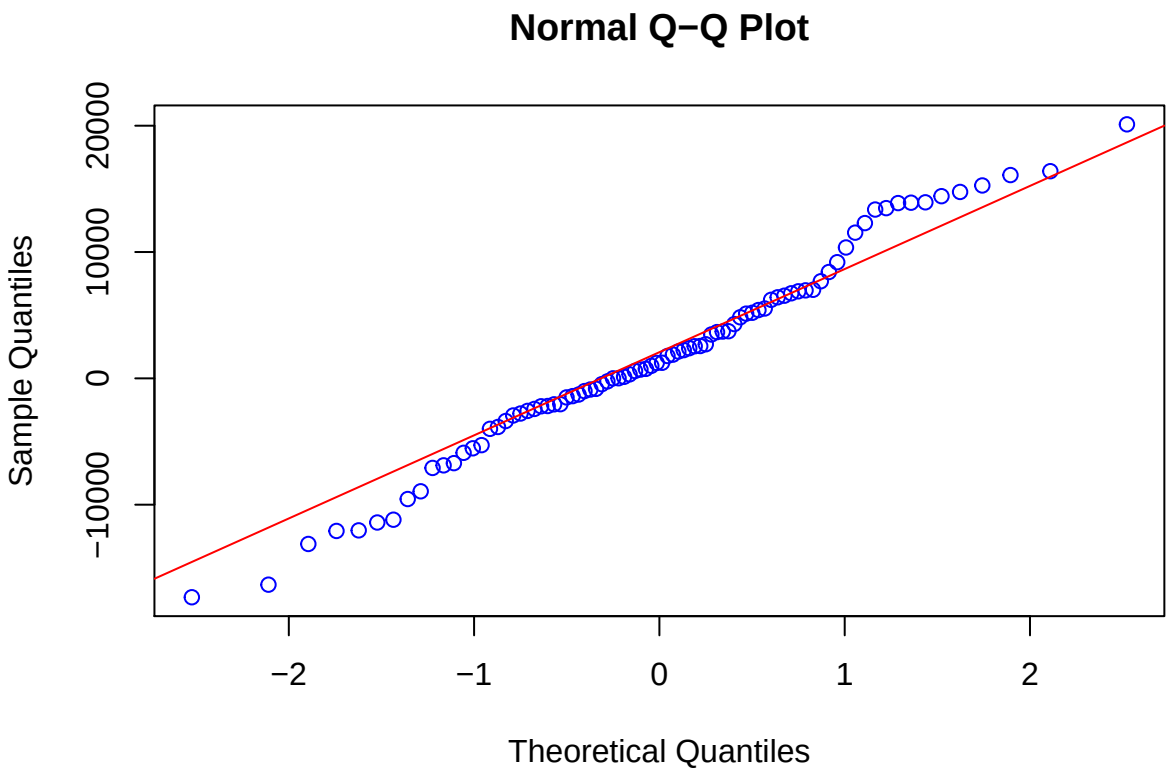

Normal Q−Q Plot
Sample Quantiles
−10000
0
10000
20000
−2
−1
0
1
2
Theoretical Quantiles

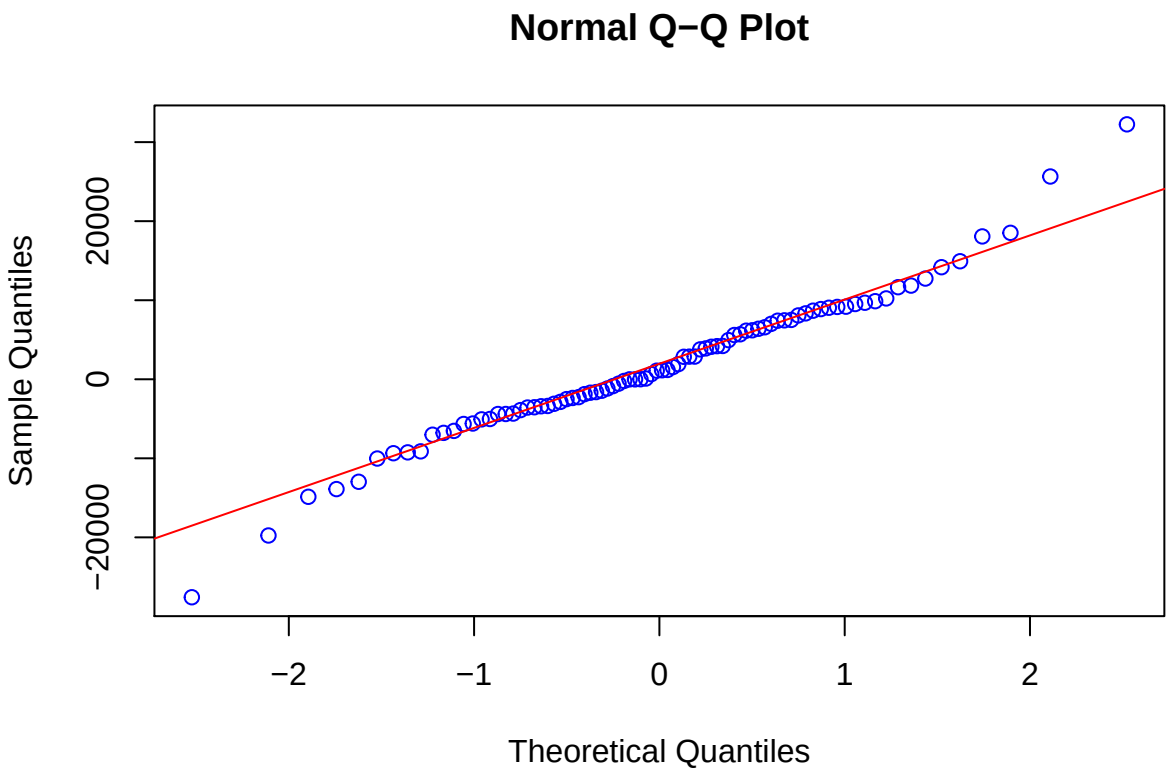

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
30000
10000
0
−10000
−2
−1
0
1
2

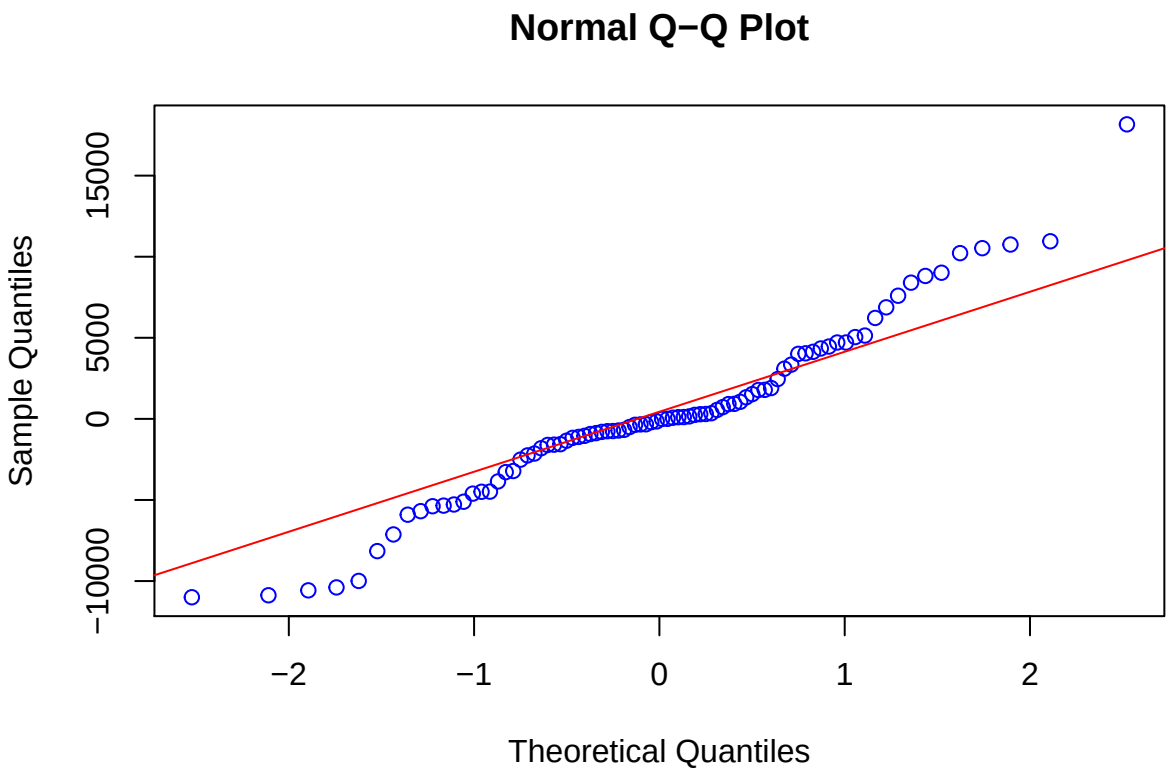

Normal Q−Q Plot
Sample Quantiles
15000
5000
0
−5000
−2
−1
0
1
2
Theoretical Quantiles

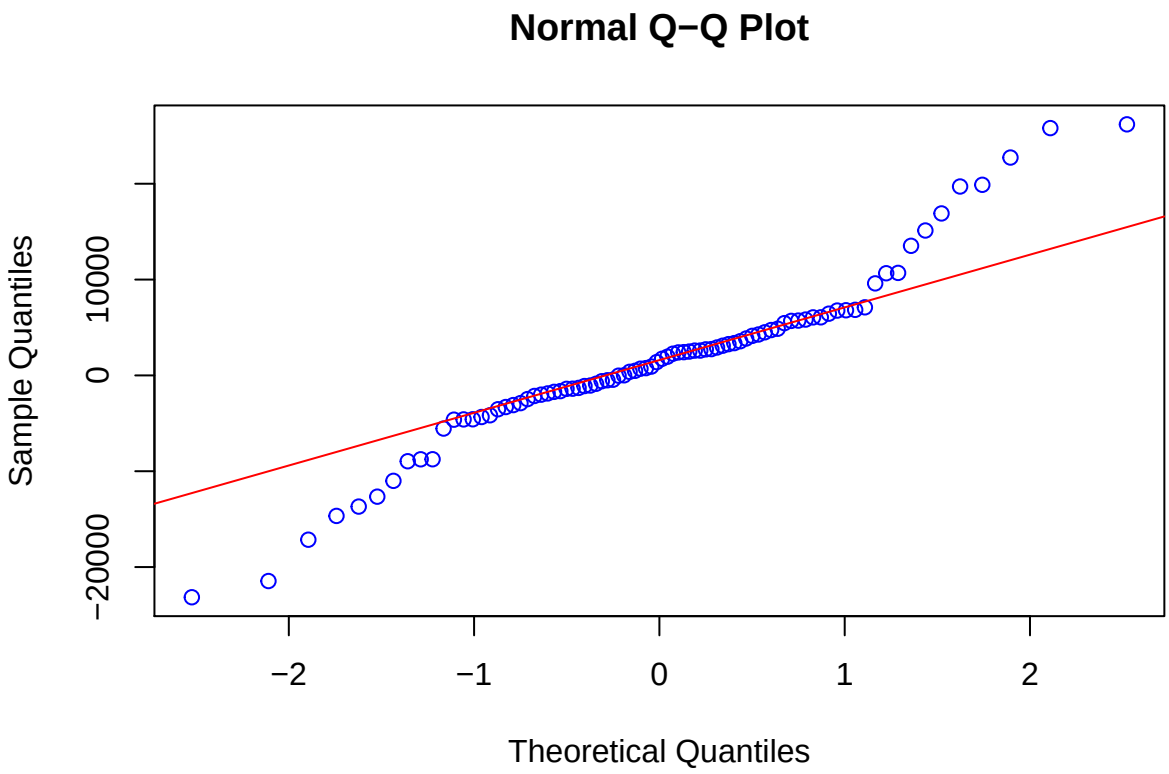

Normal Q−Q Plot
Sample Quantiles
10000
0
−20000
−2
−1
0
1
2
Theoretical Quantiles

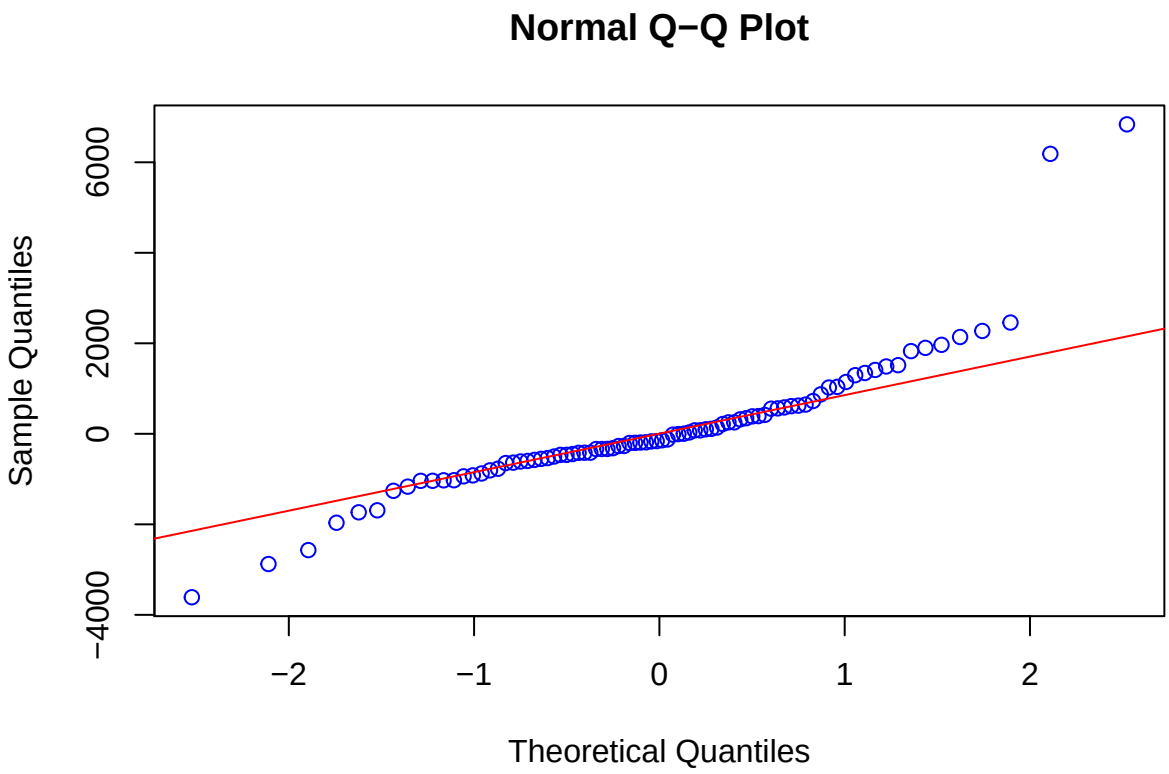

Normal Q−Q Plot
Sample Quantiles
30000
10000
−10000
−30000
−2
−1
0
1
2
Theoretical Quantiles

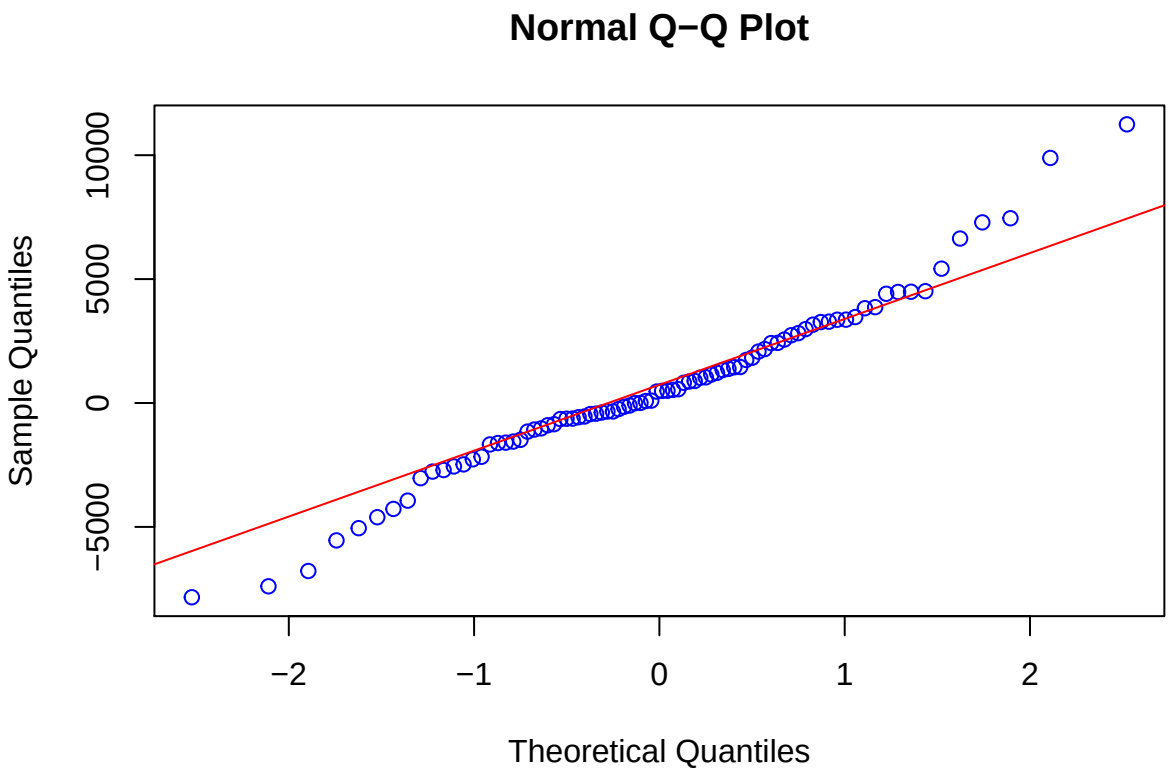

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
20000
10000
0
−10000
−2
−1
0
1
2

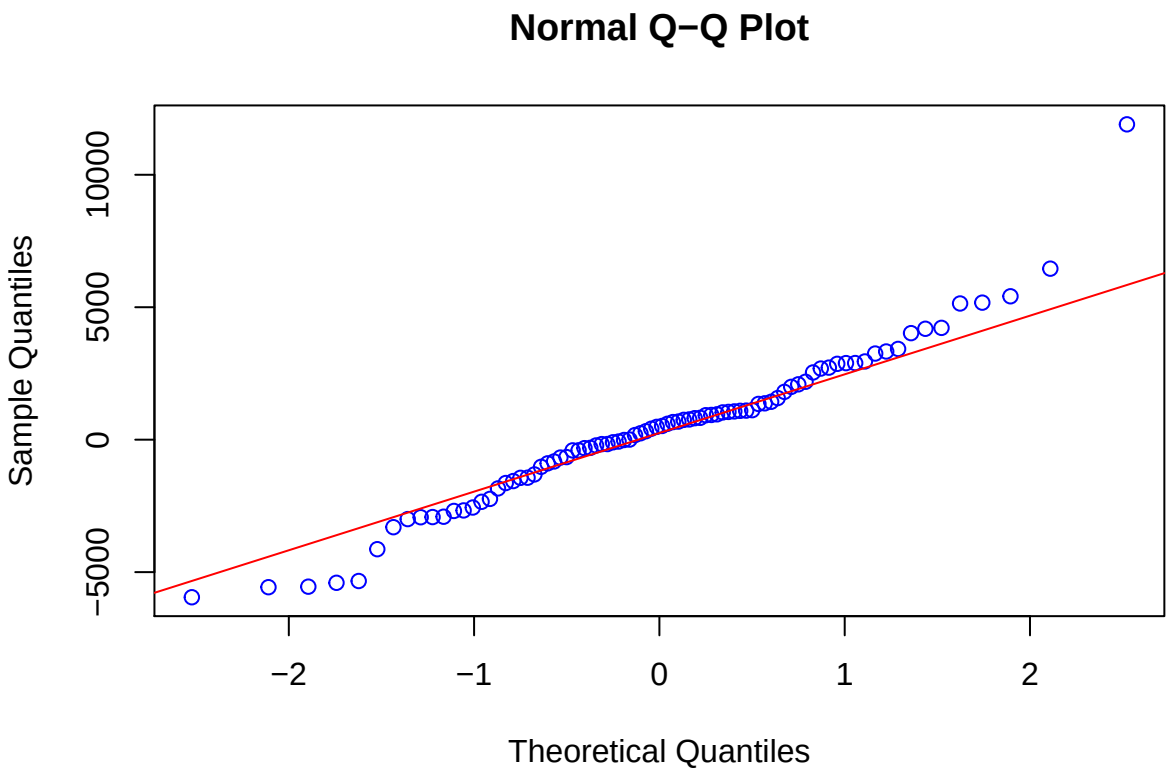

Normal Q−Q Plot
Sample Quantiles
20000
0
−20000
−2
−1
0
1
2
Theoretical Quantiles

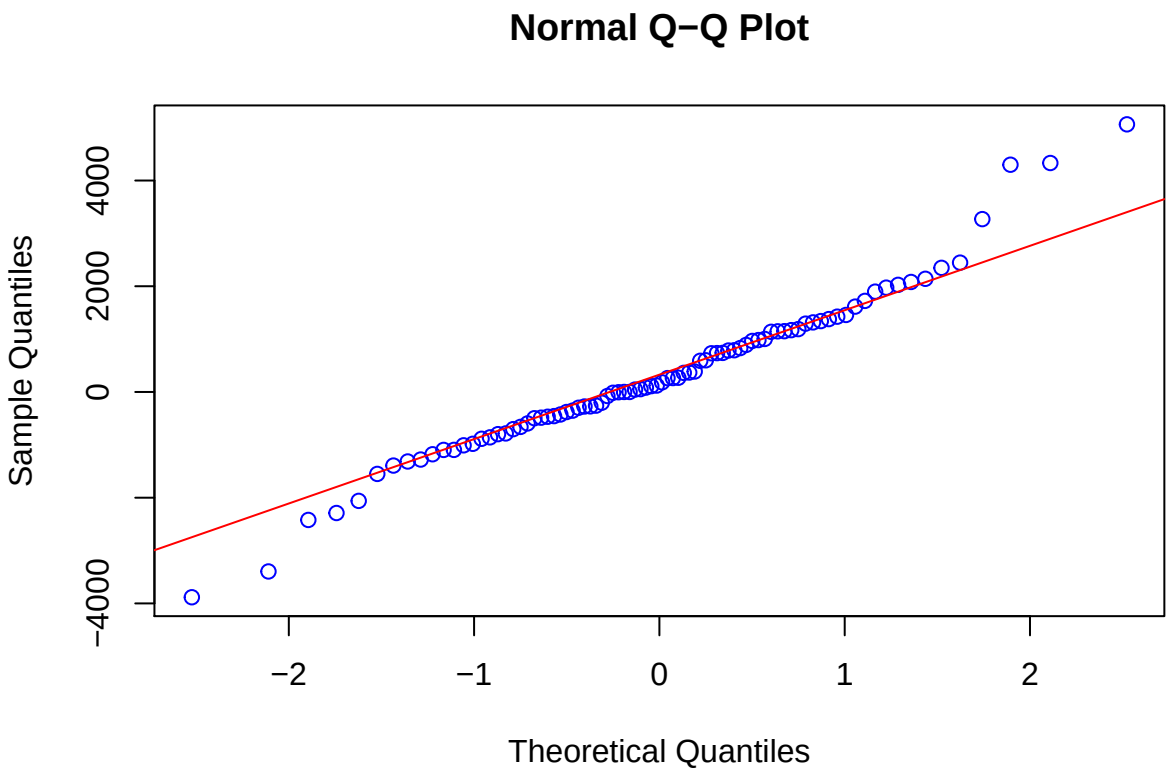

Normal Q−Q Plot
Sample Quantiles
15000
5000
0
−10000
−2
−1
0
1
2
Theoretical Quantiles

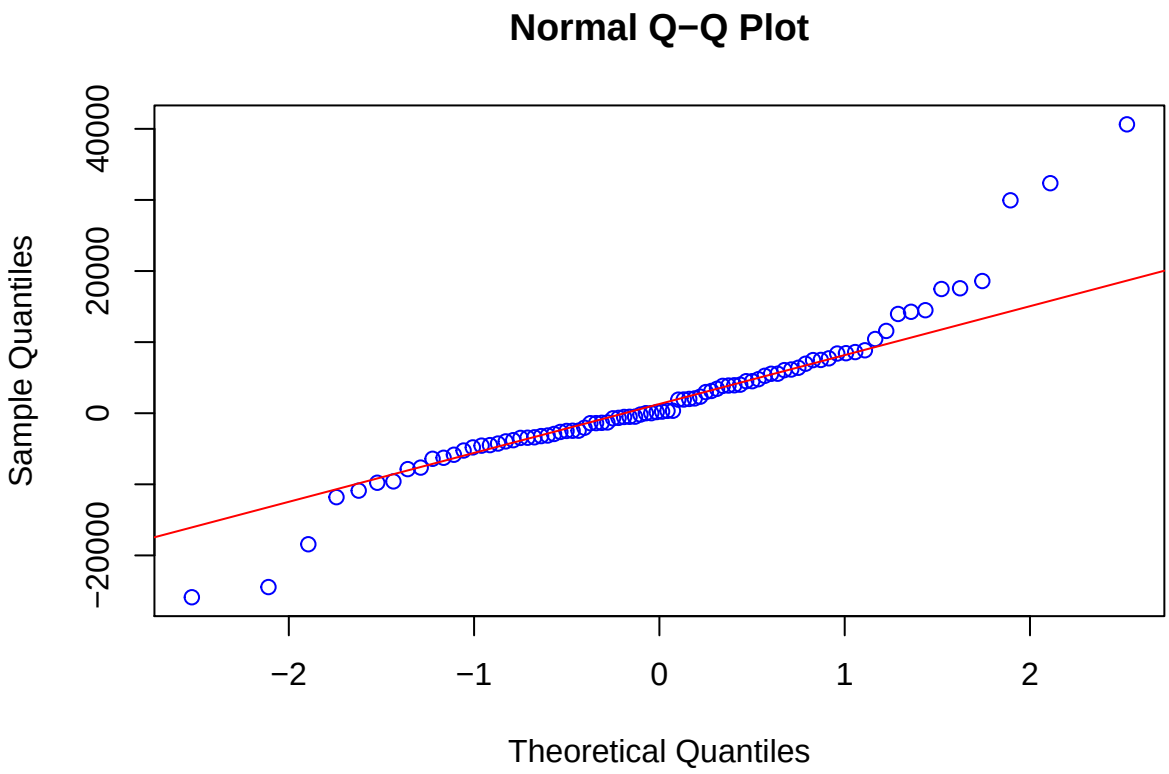

Normal Q−Q Plot
Sample Quantiles
−20000
0
10000
−2
−1
0
1
2
Theoretical Quantiles

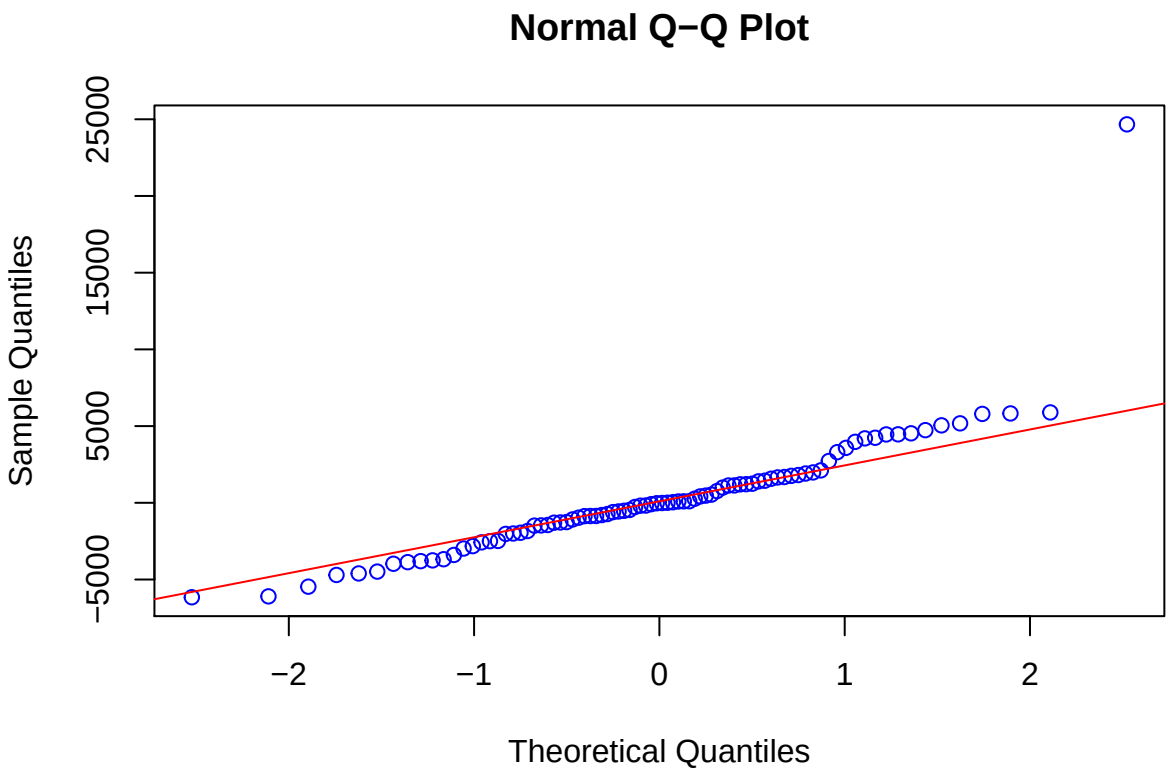

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
6000
2000
0
−4000
−2
−1
0
1
2

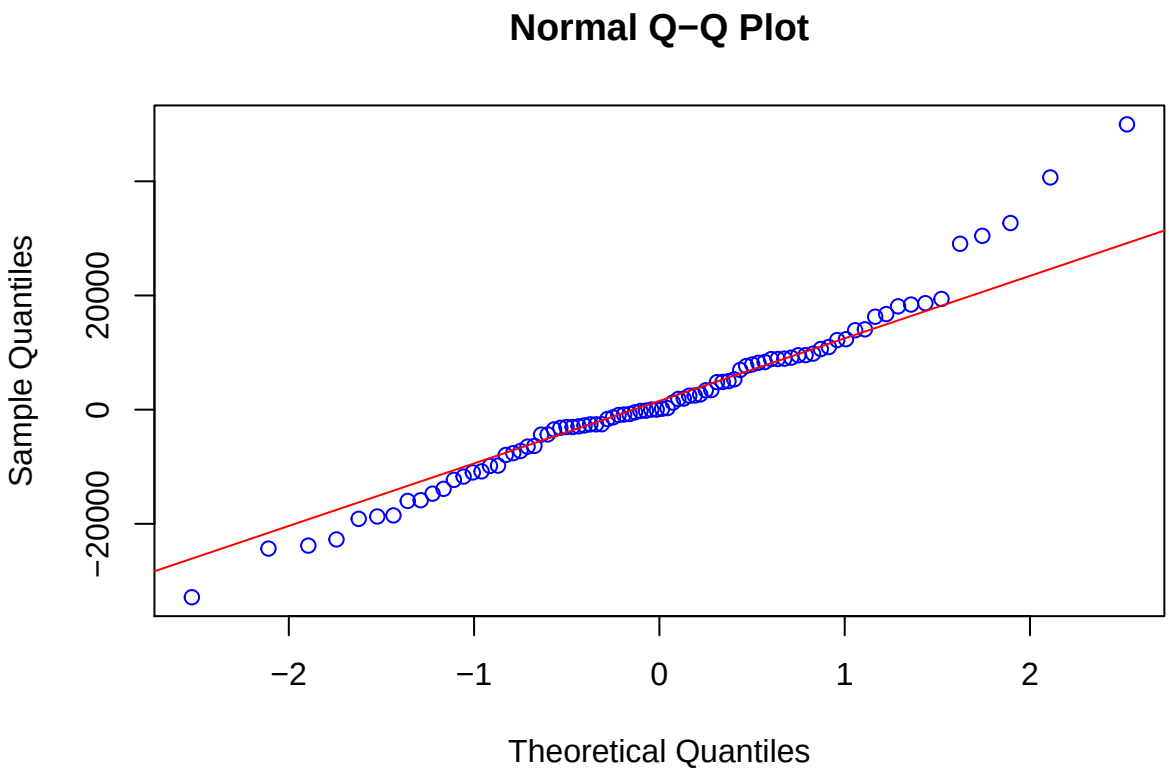

Normal Q−Q Plot
Sample Quantiles
10000
5000
0
−5000
−2
−1
0
1
2
Theoretical Quantiles

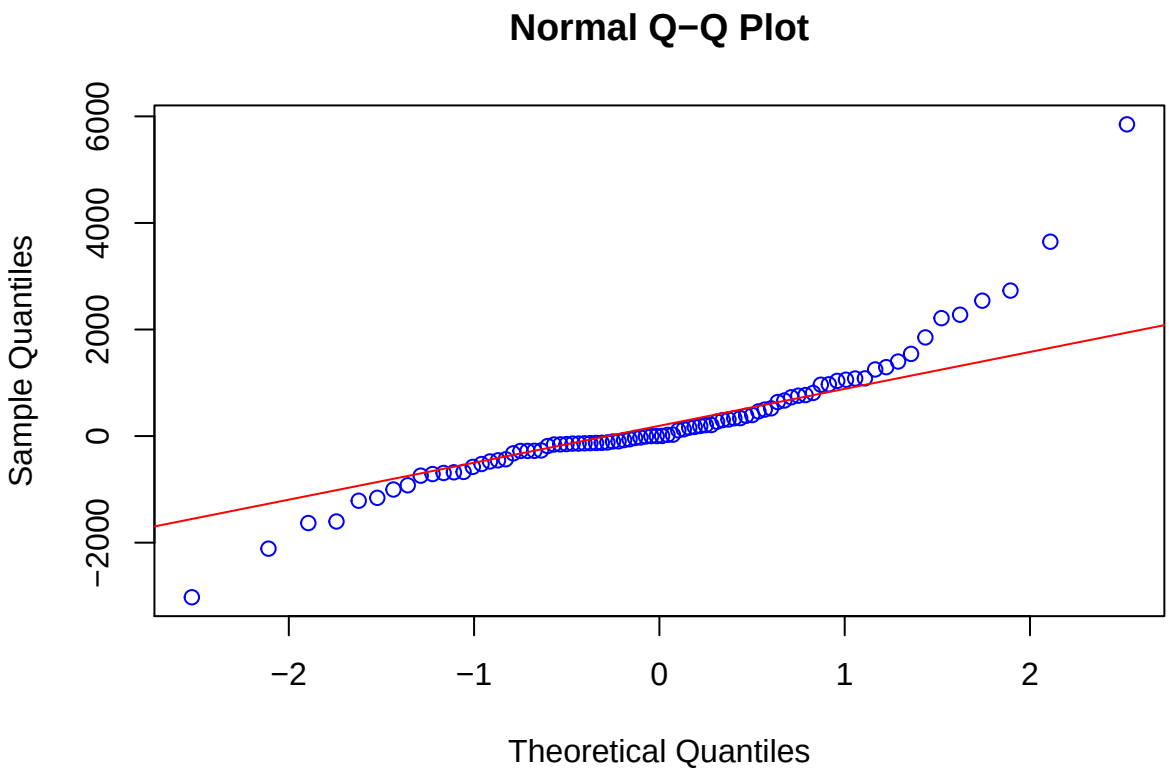

Normal Q−Q Plot
Sample Quantiles
10000
5000
0
−5000
−2
−1
0
1
2
Theoretical Quantiles

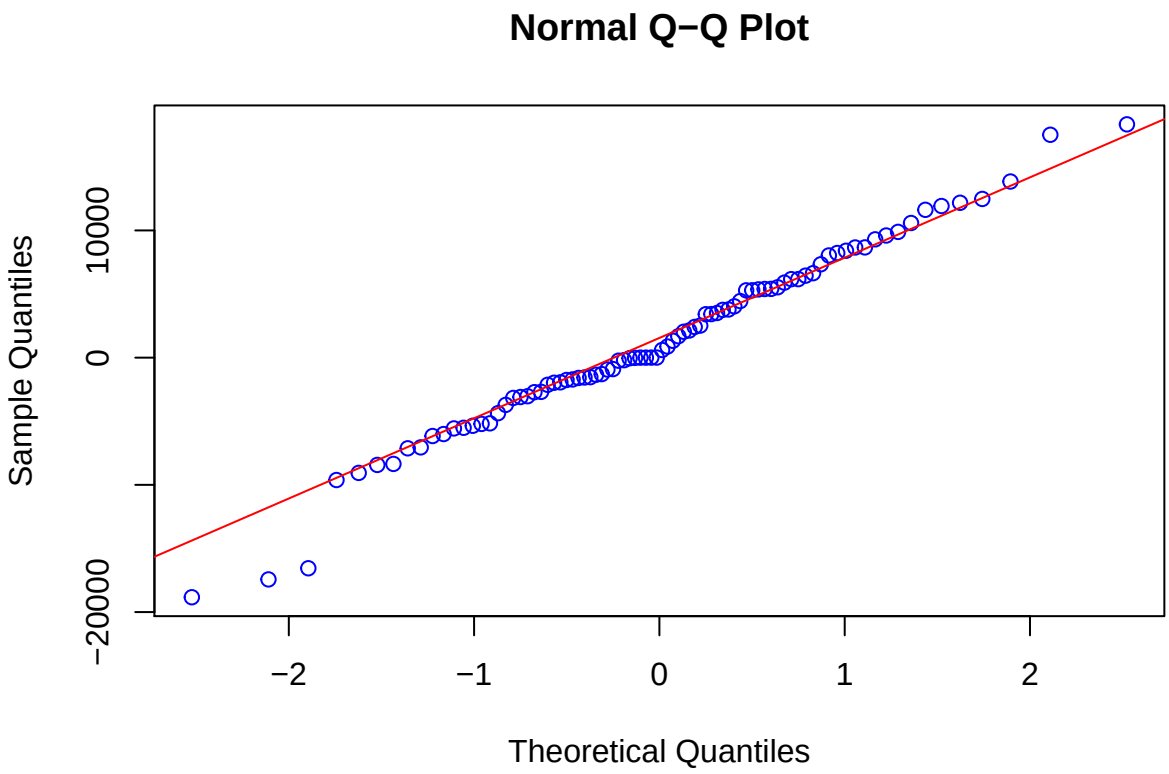

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
−4000
0
2000
4000
−2
−1
0
1
2

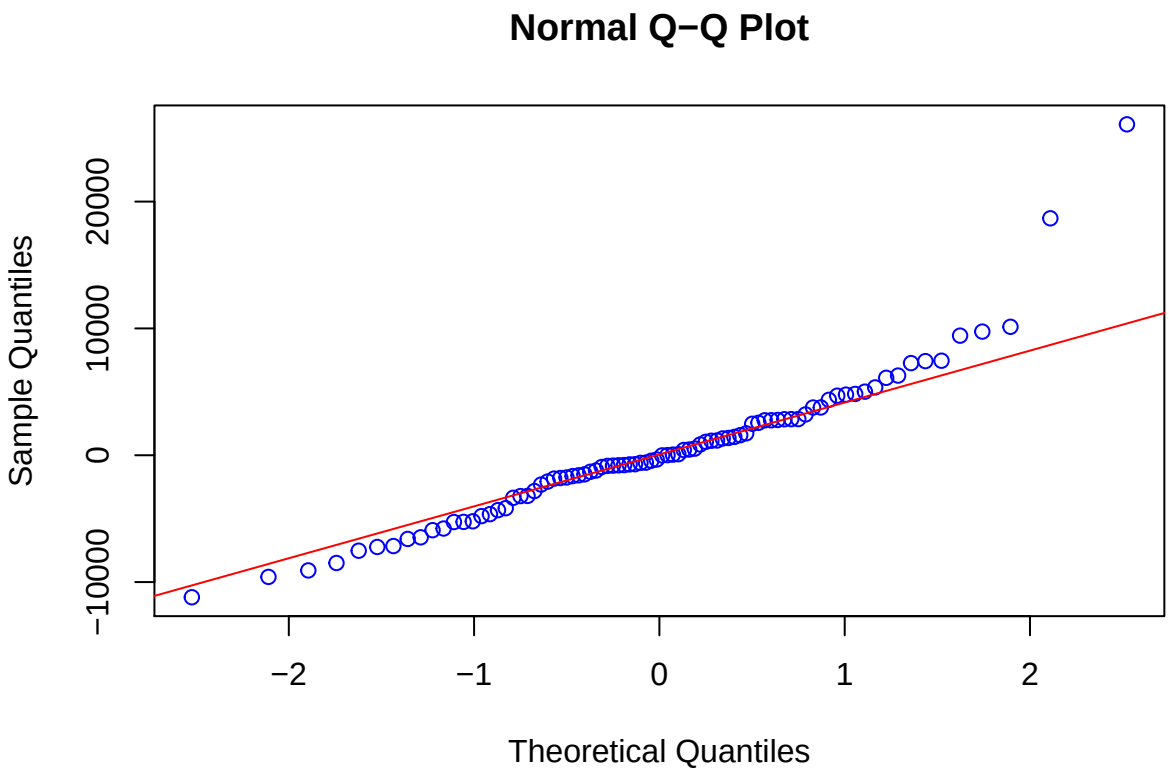

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
40000
20000
0
−20000
−2
−1
0
1
2

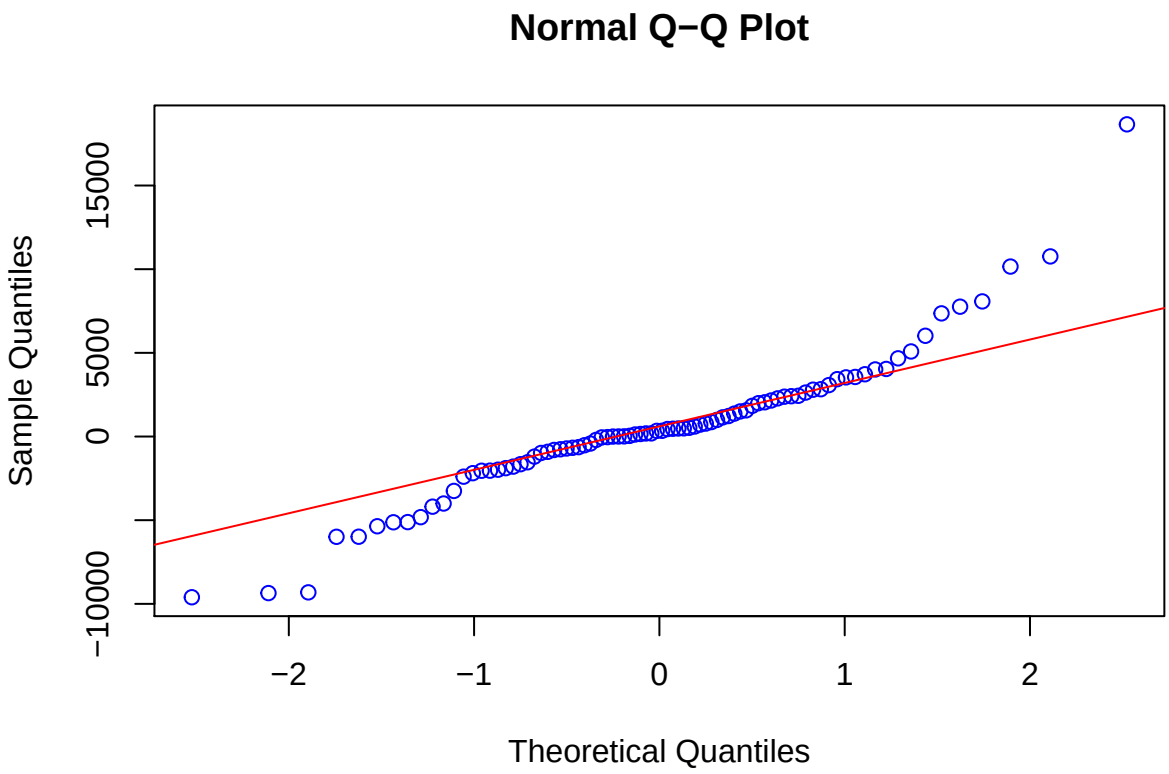

Normal Q−Q Plot
Sample Quantiles
25000
15000
5000
−5000
−2
−1
0
1
2
Theoretical Quantiles

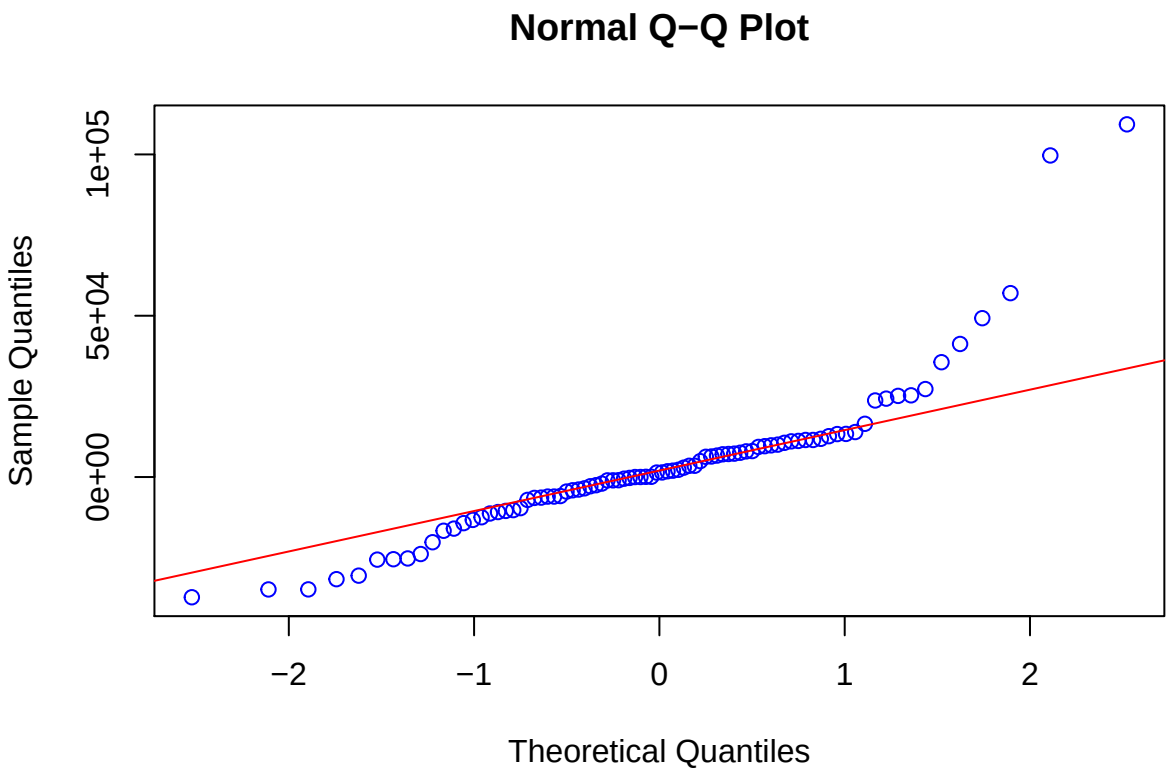

Normal Q−Q Plot
Sample Quantiles
−20000
0
20000
−2
−1
0
1
2
Theoretical Quantiles

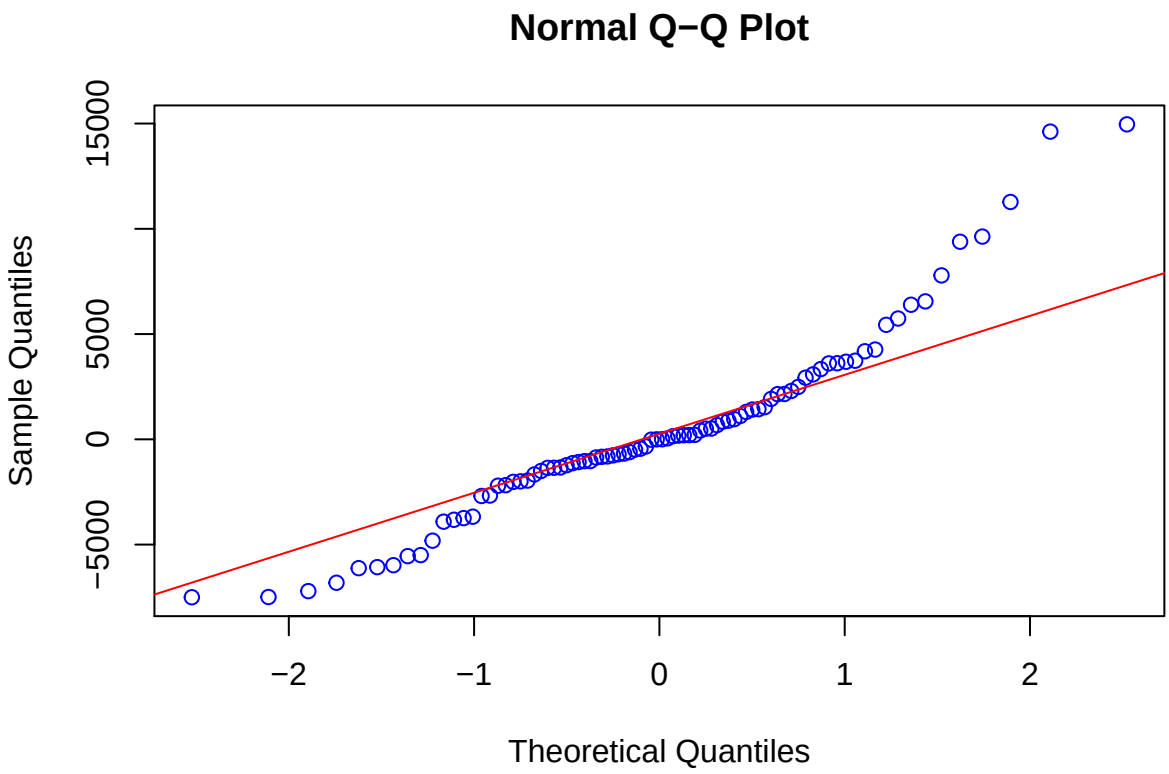

Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
6000
4000
2000
0
−2000
−2
−1
0
1
2

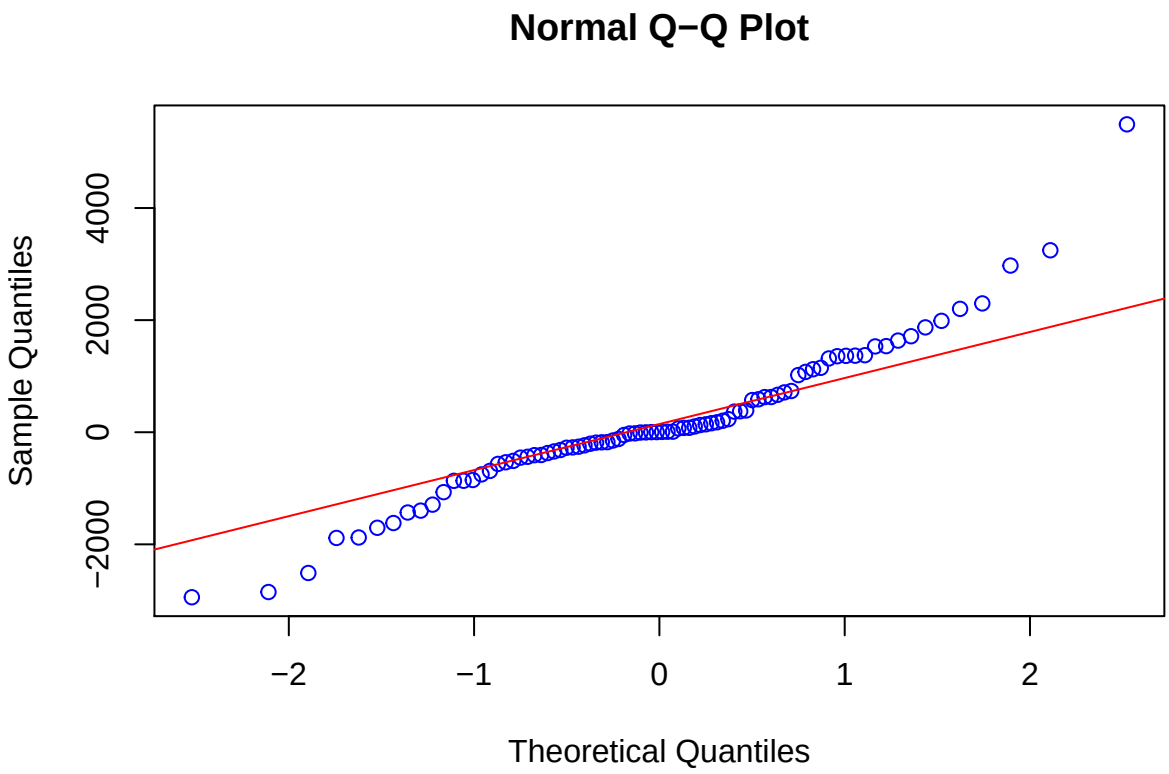

Normal Q−Q Plot
Sample Quantiles
10000
0
−20000
−2
−1
0
1
2
Theoretical Quantiles

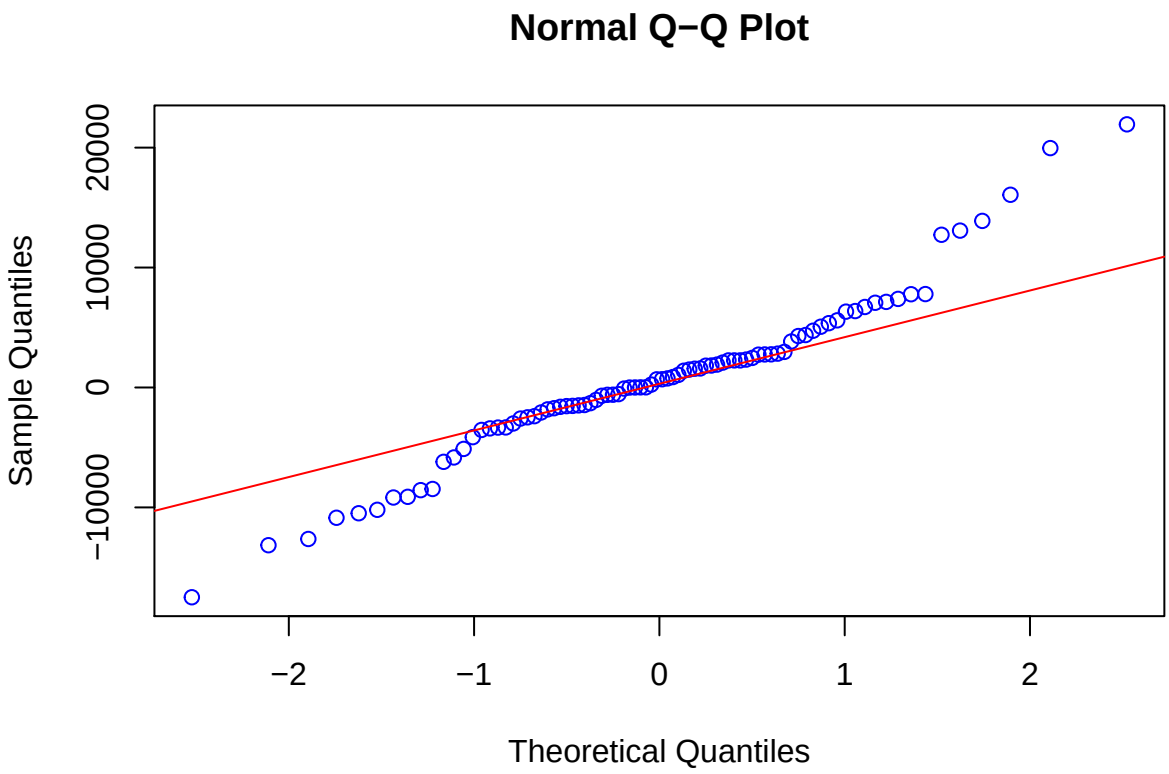

Normal Q−Q Plot
Sample Quantiles
−10000
0
10000
20000
−2
−1
0
1
2
Theoretical Quantiles

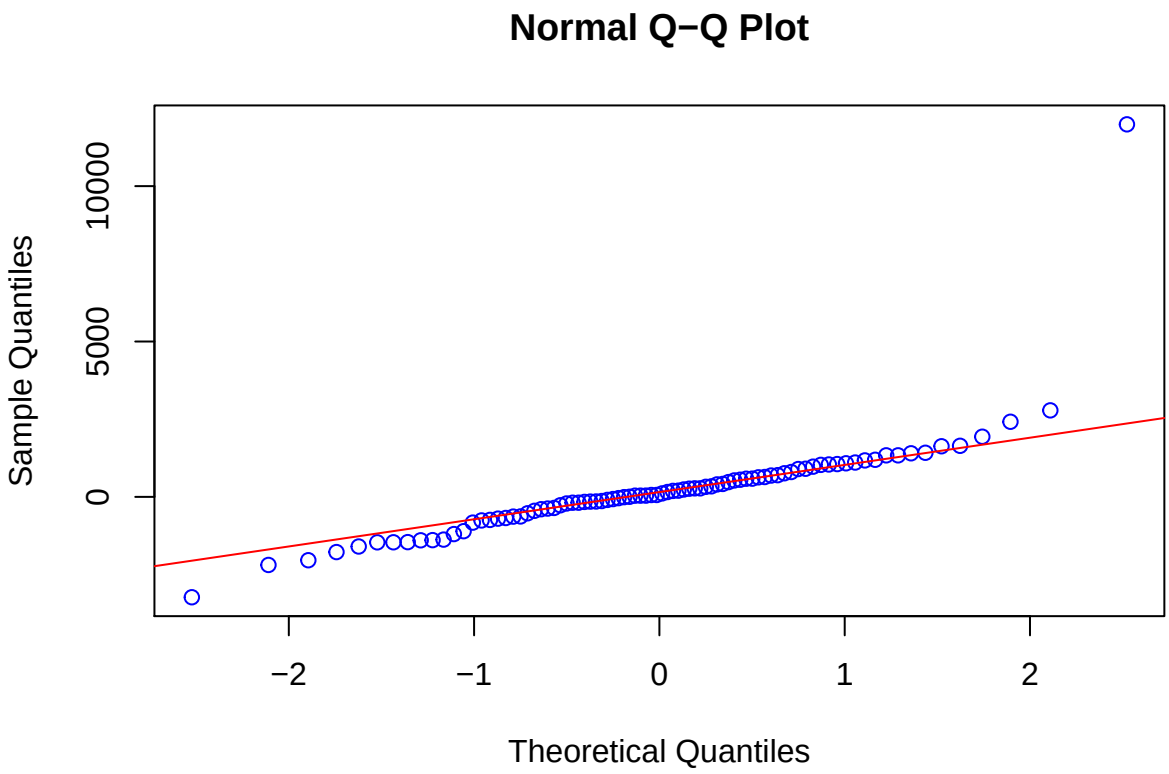

Normal Q−Q Plot
Sample Quantiles
15000
5000
0
−10000
−2
−1
0
1
2
Theoretical Quantiles

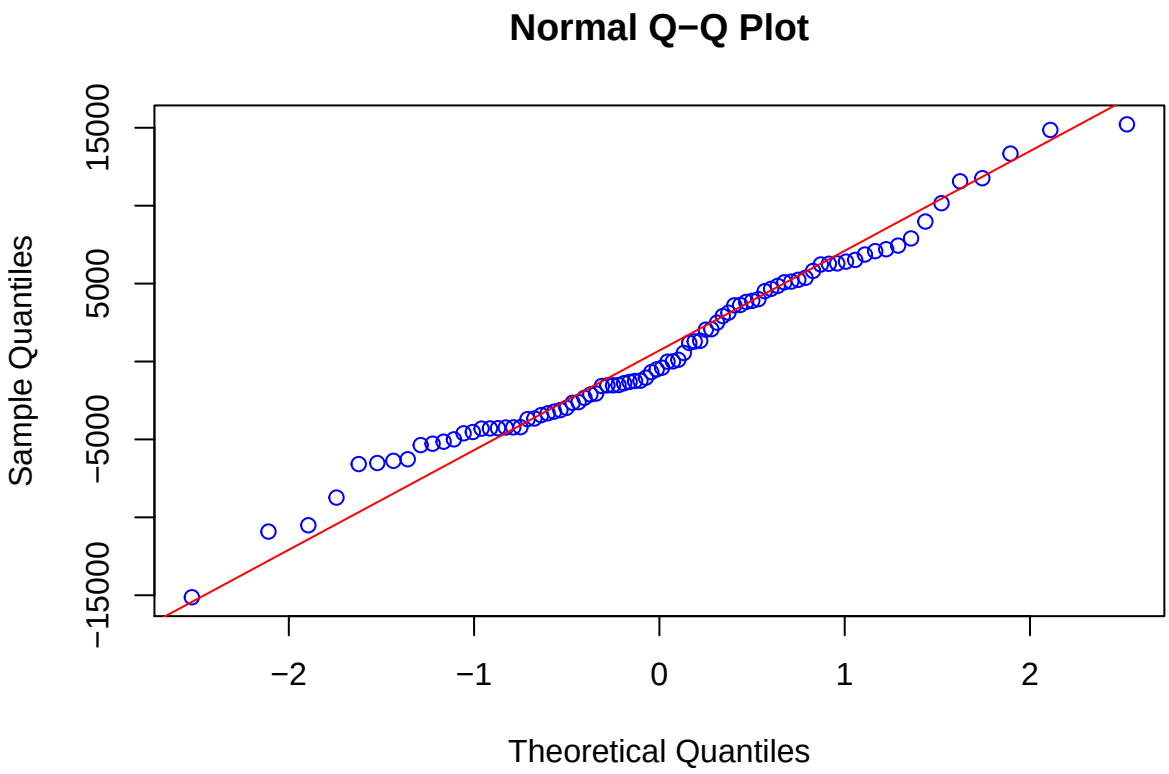

Normal Q−Q Plot
Sample Quantiles
0e+00
5e+04
1e+05
−2
−1
0
1
2
Theoretical Quantiles

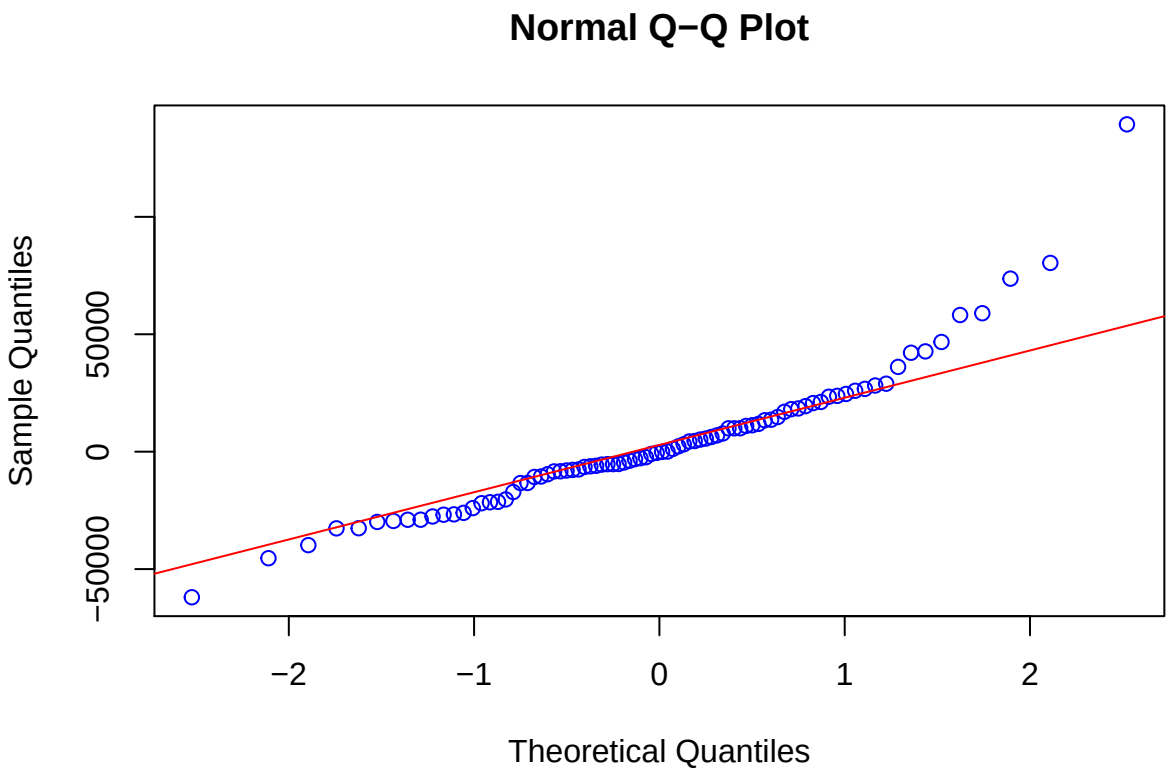

Normal Q−Q Plot
Sample Quantiles
15000
5000
0
−5000
−2
−1
0
1
2
Theoretical Quantiles

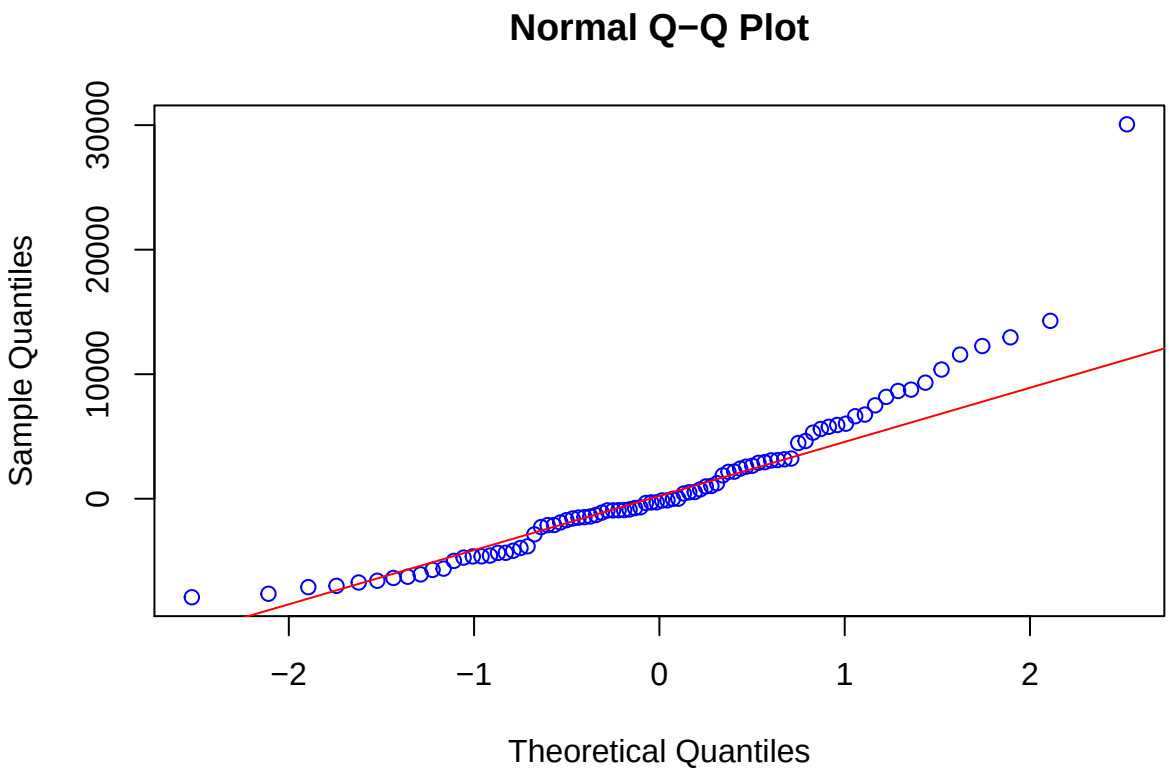

Normal Q−Q Plot
Sample Quantiles
−2000
0
2000
4000
−2
−1
0
1
2
Theoretical Quantiles

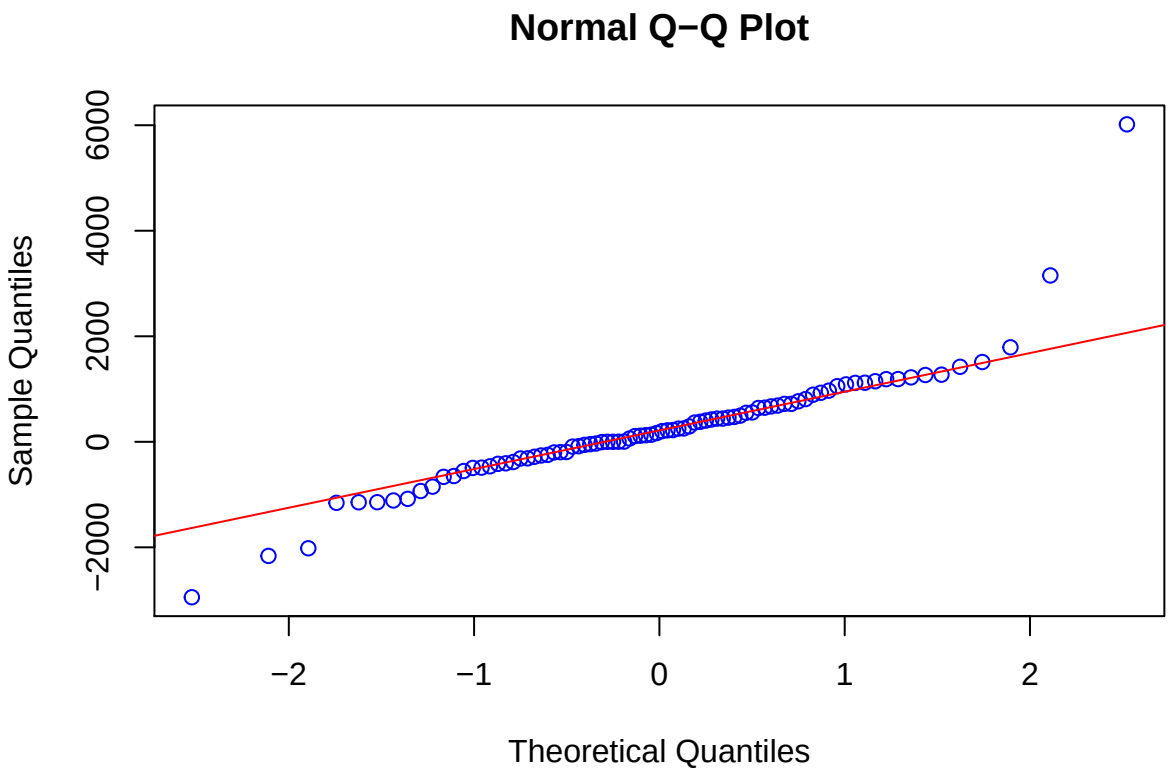

Normal Q−Q Plot
Sample Quantiles
−10000
0
10000
20000
−2
−1
0
1
2
Theoretical Quantiles

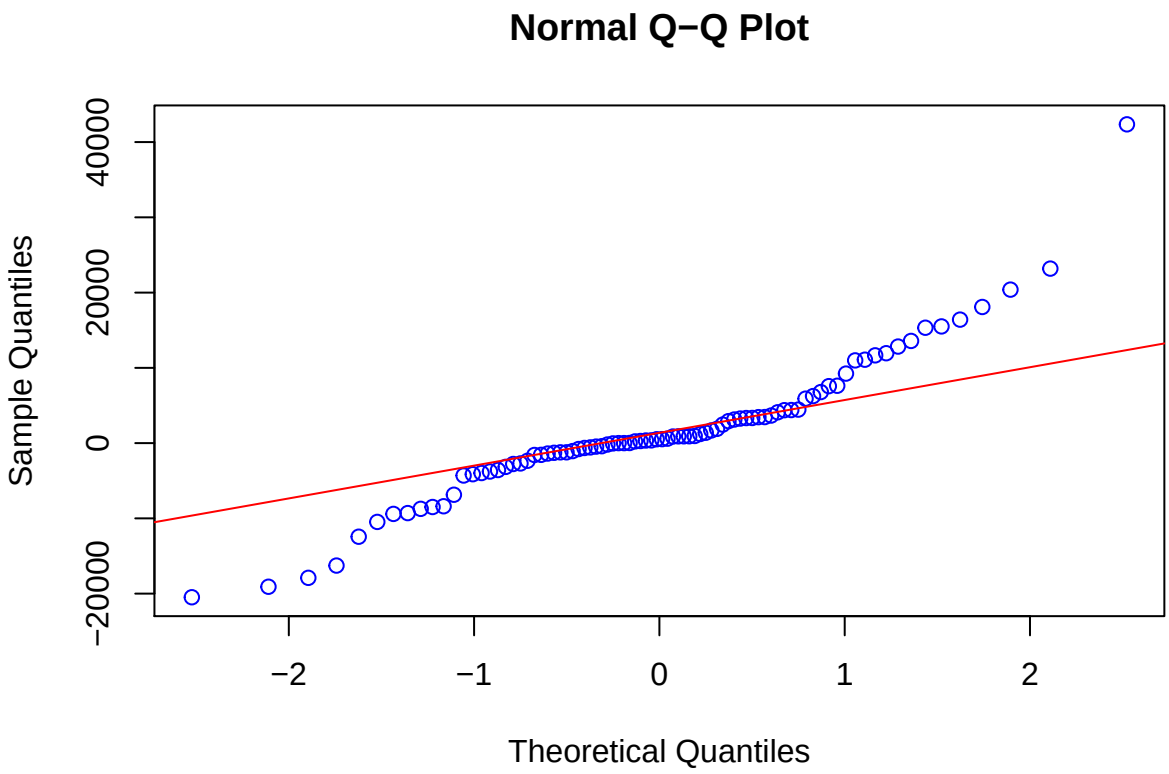

Normal Q−Q Plot
Sample Quantiles
0
5000
10000
−2
−1
0
1
2
Theoretical Quantiles

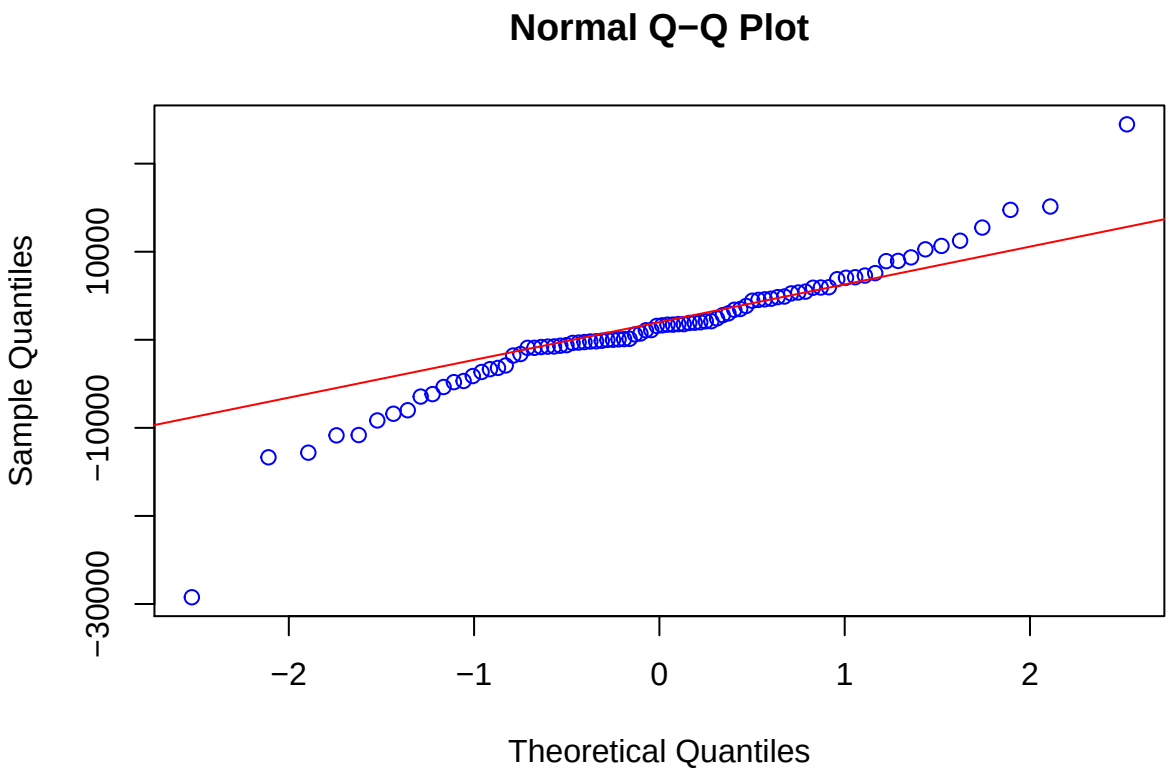

Normal Q−Q Plot
Sample Quantiles
15000
5000
−5000
−15000
−2
−1
0
1
2
Theoretical Quantiles

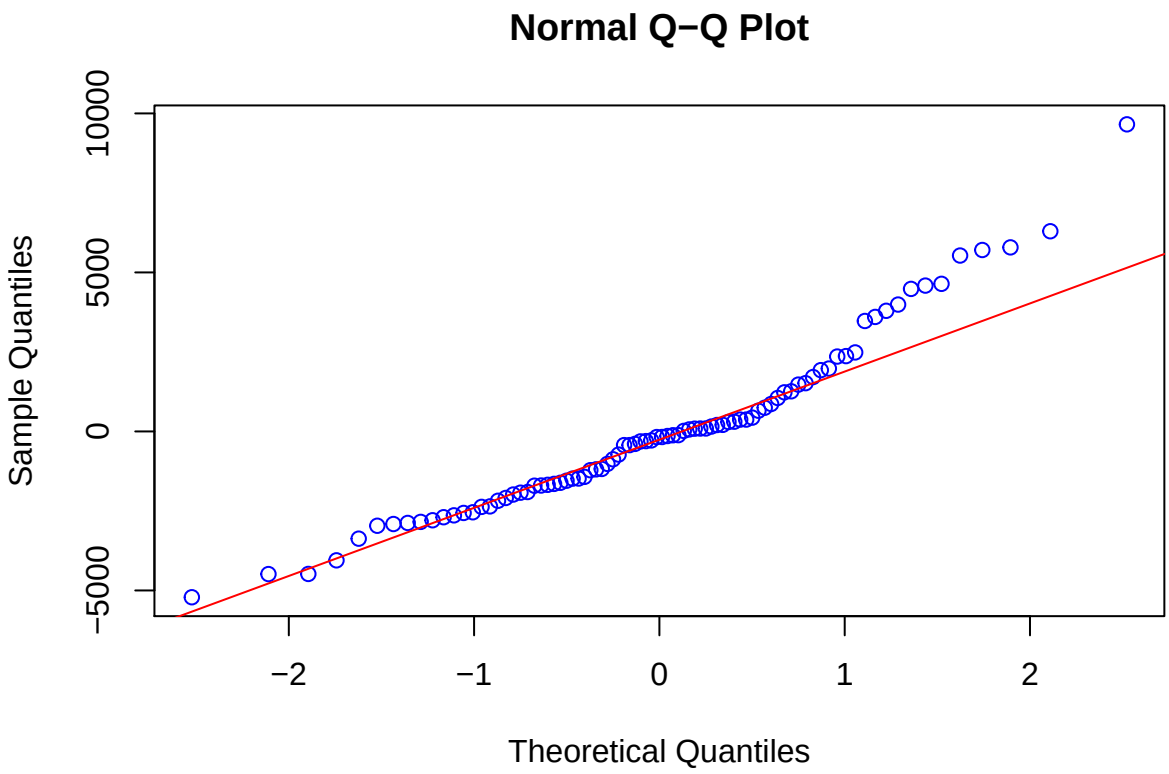

Normal Q−Q Plot
Sample Quantiles
−50000
0
50000
−2
−1
0
1
2
Theoretical Quantiles

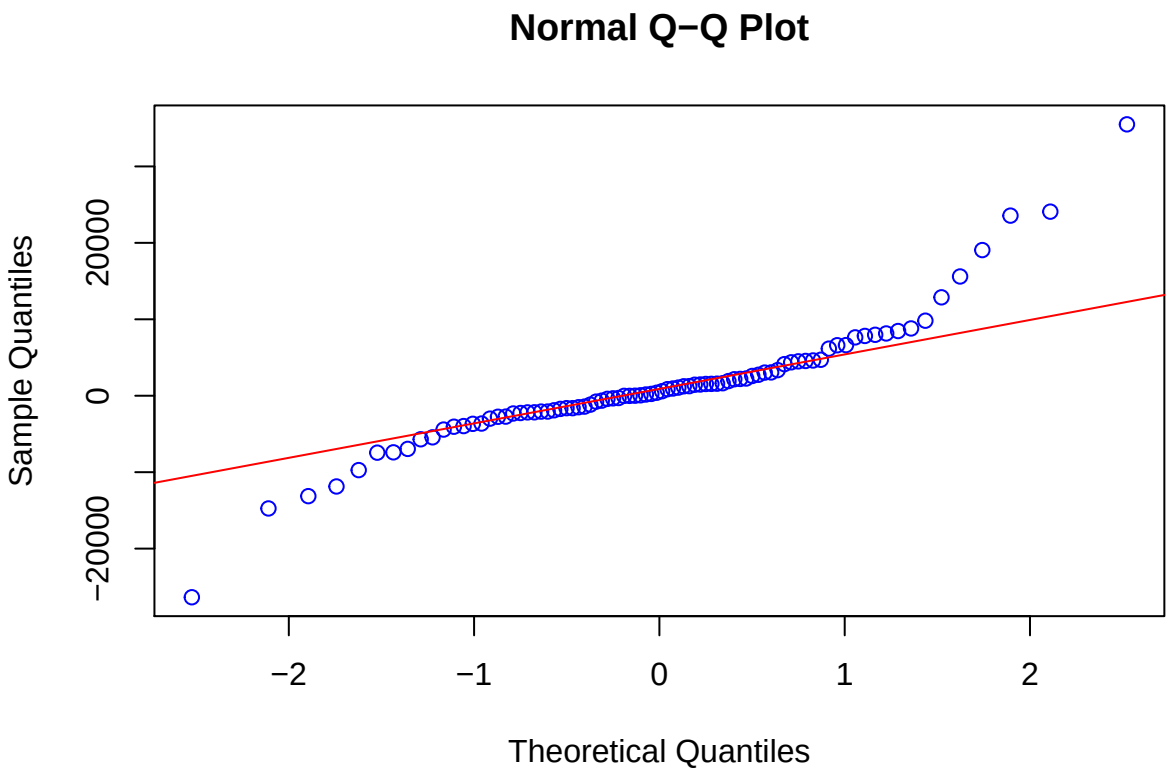

Normal Q−Q Plot
Sample Quantiles
0
10000
20000
30000
−2
−1
0
1
2
Theoretical Quantiles

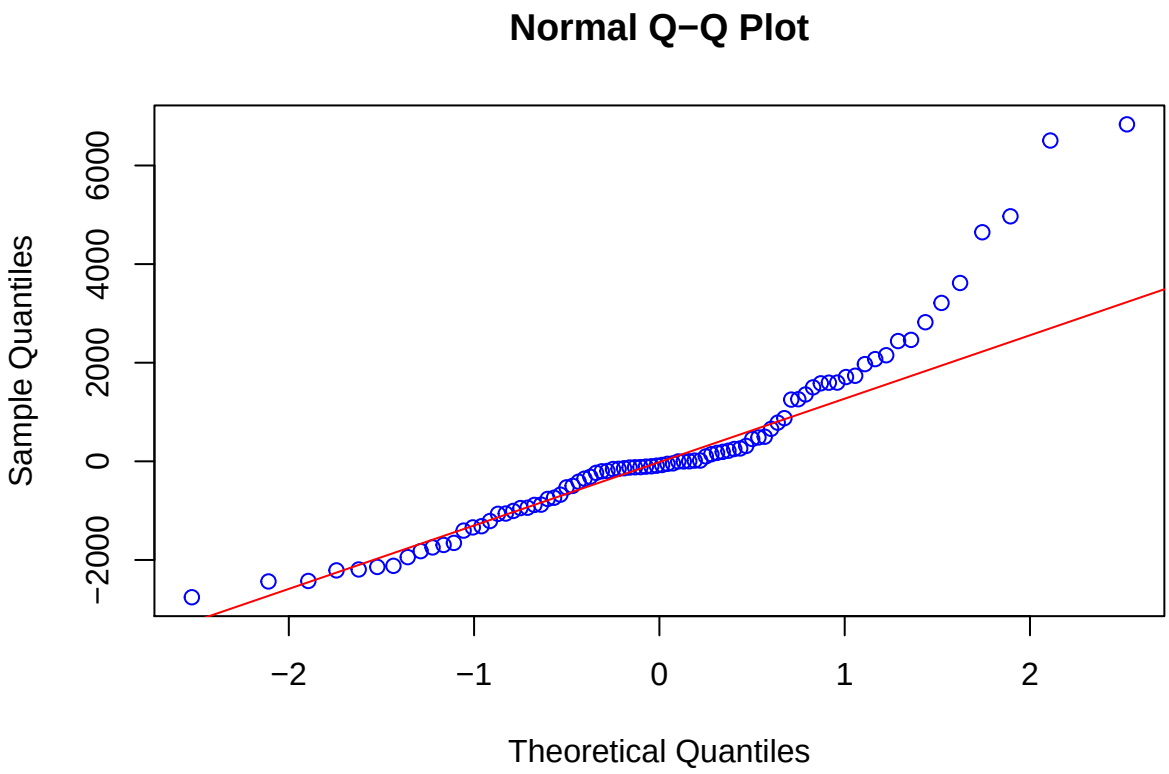

Normal Q−Q Plot
Sample Quantiles
−2000
0
2000
4000
6000
−2
−1
0
1
2
Theoretical Quantiles


Normal Q−Q Plot
Sample Quantiles
Theoretical Quantiles
40000
20000
0
−20000
−2
−1
0
1
2


Normal Q−Q Plot
Sample Quantiles
10000
−10000
−30000
−2
−1
0
1
2
Theoretical Quantiles


Normal Q−Q Plot
Sample Quantiles
10000
5000
0
−5000
−2
−1
0
1
2
Theoretical Quantiles


Normal Q−Q Plot
Sample Quantiles
20000
0
−20000
−2
−1
0
1
2
Theoretical Quantiles

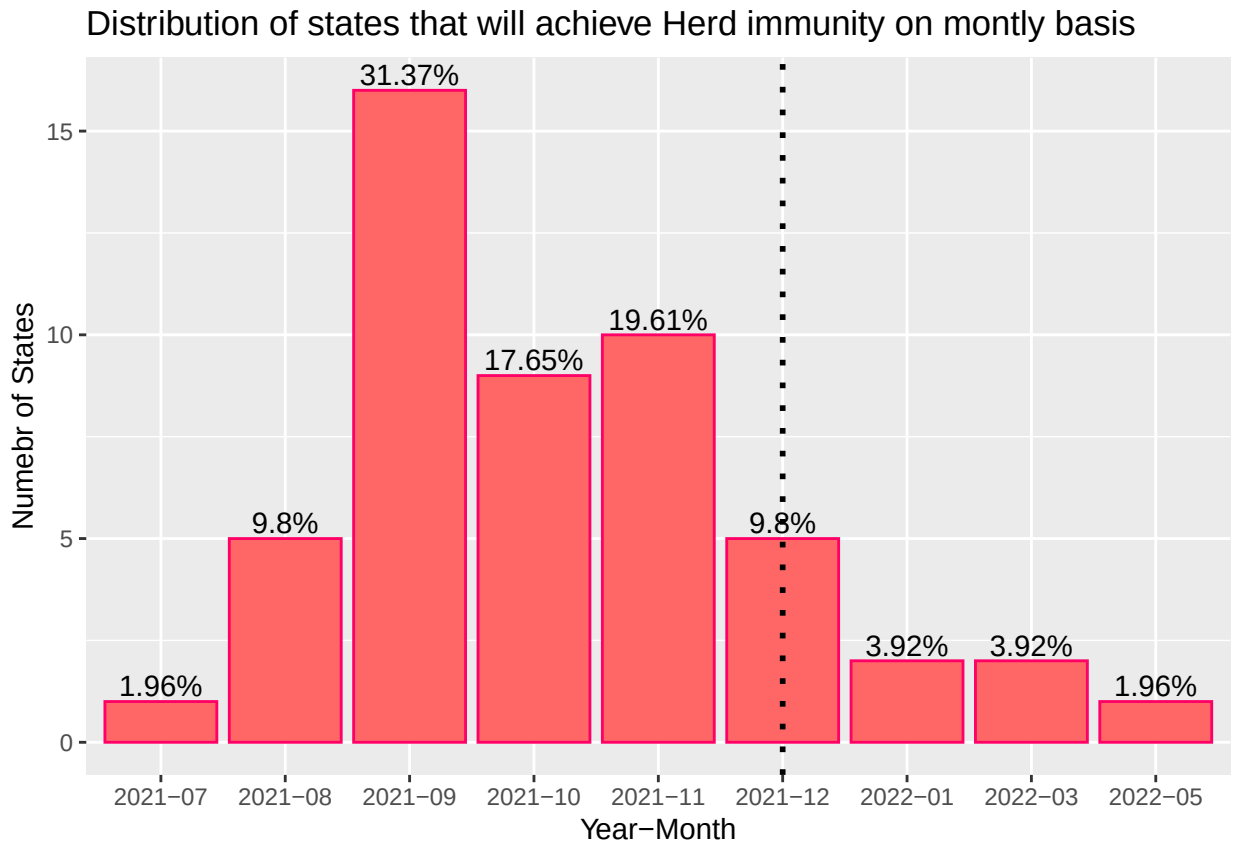



```
output
```

```
##                  states      herd_immunity
## 1               Alabama   15,January, 2022
## 2                Alaska   11,October, 2021
## 3               Arizona  24,November, 2021
## 4              Arkansas  16,December, 2021
## 5            California   15,October, 2021
## 6              Colorado  08,November, 2021
## 7           Connecticut    28,August, 2021
## 8              Delaware  11,November, 2021
## 9  District of Columbia       10,May, 2022
## 10              Florida  19,November, 2021
## 11              Georgia     11,March, 2022
## 12               Hawaii 29,September, 2021
## 13                Idaho  17,November, 2021
## 14             Illinois  15,November, 2021
## 15              Indiana   30,October, 2021
## 16                 Iowa   23,October, 2021
## 17               Kansas 16,September, 2021
## 18             Kentucky 25,September, 2021
## 19            Louisiana  23,November, 2021
## 20                Maine    21,August, 2021
## 21             Maryland    30,August, 2021
## 22        Massachusetts 07,September, 2021
## 23             Michigan 21,September, 2021
```

```
## 24            Minnesota 11,September, 2021
## 25          Mississippi  16,December, 2021
## 26             Missouri  01,November, 2021
## 27              Montana   24,October, 2021
## 28             Nebraska 29,September, 2021
## 29               Nevada   22,October, 2021
## 30        New Hampshire 13,September, 2021
## 31           New Jersey    15,August, 2021
## 32           New Mexico 22,September, 2021
## 33       North Carolina  11,December, 2021
## 34         North Dakota 03,September, 2021
## 35                 Ohio 30,September, 2021
## 36             Oklahoma 10,September, 2021
## 37               Oregon   07,October, 2021
## 38         Pennsylvania  01,November, 2021
## 39          Puerto Rico     06,March, 2022
## 40         Rhode Island      31,July, 2021
## 41       South Carolina  20,December, 2021
## 42         South Dakota 11,September, 2021
## 43            Tennessee  08,December, 2021
## 44                Texas   06,October, 2021
## 45                 Utah  27,November, 2021
## 46              Vermont    22,August, 2021
## 47             Virginia 25,September, 2021
## 48           Washington 23,September, 2021
## 49        West Virginia   05,January, 2022
## 50            Wisconsin 07,September, 2021
## 51              Wyoming   05,October, 2021
```

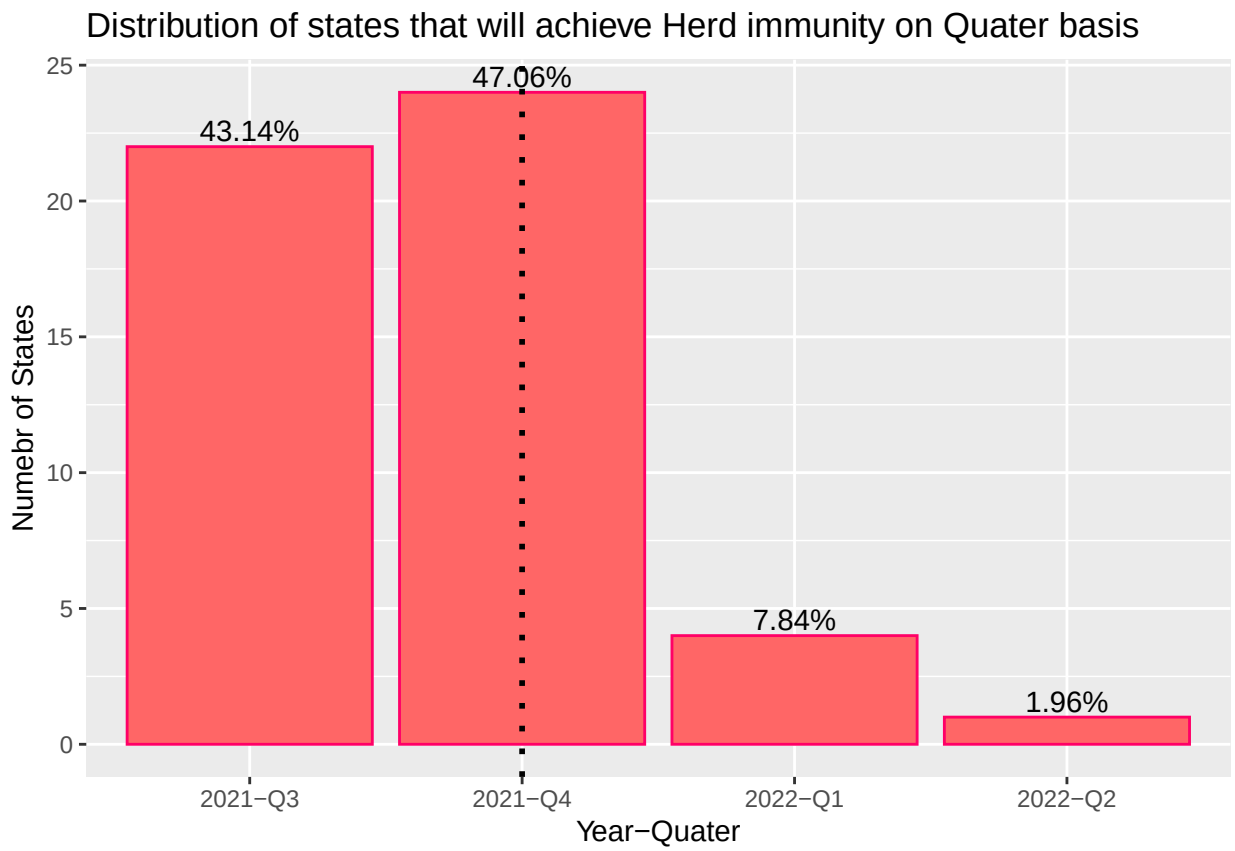
```r
output$herd_immunity <- as.Date(output$herd_immunity, format = "%d,%B, %Y")
output$monthYear <- format(output$herd_immunity , "%Y-%m")
output$quaterYear <- paste(format(output$herd_immunity,"%Y"),
                                quarters(output$herd_immunity), sep="-")

output %>%
  group_by(monthYear) %>%
  summarise(count = n()) %>%
  ggplot(aes(x = monthYear, y = count))+
  geom_bar(stat = 'identity', color="#FF0066", fill="#FF6666")+
  geom_text(aes(label= paste(
            round(count/ length(states)*100,2),
            "%",
            sep =""
          )),
        position=position_dodge(width=0.9), vjust=-0.25)+
  geom_vline(xintercept = '2021-12',
             linetype="dotted",
             size = 1)+
  labs(x = "Year-Month",
       y = "Numebr of States",
       title = "Distribution of states that will achieve Herd immunity on montly basis")
```




```
ggsave("monthly.png")
```

```
## Saving 6.5 x 4.5 in image
```

```
output %>%
  group_by(quaterYear) %>%
  summarise(count = n()) %>%
  ggplot(aes(x = quaterYear, y = count))+
  geom_bar(stat = 'identity', color="#FF0066", fill="#FF6666")+
  geom_text(aes(label= paste(
            round(count/ length(states)*100,2),
            "%",
            sep =""
          )),
        position=position_dodge(width=0.9), vjust=-0.25)+
  geom_vline(xintercept = '2021-Q4',
             linetype="dotted",
             size = 1)+
  labs(x = "Year-Quater",
       y = "Numebr of States",
       title = "Distribution of states that will achieve Herd immunity on Quater basis")
```




```
ggsave("quater.png")
```

```
## Saving 6.5 x 4.5 in image
```