Homework4Part1

Rafael Espinosa Castaneda

2023-02-18

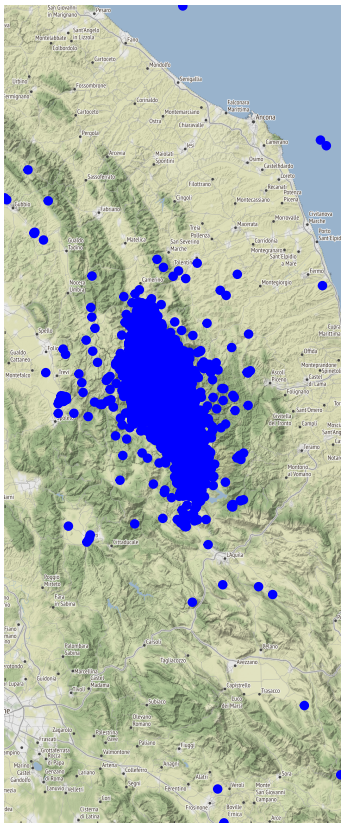
```
# --------------------------Setting the parameters to be used-------------------------


library(tidyverse)
library(ggmap)
library(tibble)
library(spatstat)
library(sf)

earthquakes_italy<-read_csv("/Users/rafa/Documents/Master Austin/MAESTRIA_AUSTIN/Advanced Predictive Mo
```

Part a

```
min_lat <-100000
max_lat<-0


min_lon <-100000
max_lon<-0


for (i in 1:length(earthquakes_italy$Latitude))
{
  if (earthquakes_italy$Latitude[i]<min_lat)
  {
    min_lat<-earthquakes_italy$Latitude[i]
  }

  if (earthquakes_italy$Latitude[i]>max_lat)
  {
    max_lat<-earthquakes_italy$Latitude[i]
  }


  if (earthquakes_italy$Longitude[i]<min_lon)
  {
    min_lon<-earthquakes_italy$Longitude[i]
  }

  if (earthquakes_italy$Longitude[i]>max_lon)
  {
    max_lon<-earthquakes_italy$Longitude[i]
```

```
  }

}

#left = 12.620544, bottom = 42.602125, right = 13.599014, top = 42.974511

bbox <- c(left = min_lon, bottom = min_lat, right = max_lon, top = max_lat)

ggmap(get_stamenmap(bbox, zoom =10), extent = "device")+
  geom_point(data = earthquakes_italy,
             aes(x = Longitude, y = Latitude),
             color = "blue",
             size = 1) +
  ggtitle("Earthquake locations in the center of Italy") +
  labs (x = "longitud", y = "latitud")
```

Earthquake locations in the center of Italy



ANSWER

The dataset contains data about the earthquake locations that took place in the center of Italy between August and November 2016. The data set consists of time,Latitude,Longitude,Depth/Km,Magnitude of the earthquakes. The original data set was collected from the Italian Earthquakes National Center. Links to the data set is

https://www.kaggle.com/datasets/blackecho/italy-earthquakes and the site where the dataset was collected from is http://cnt.rm.ingv.it/

Part b

ANSWER

The data is a collection of locations at the center of Italy where earthquakes took place; with covariate and mark the depth and time of the earthquakes respectively (time is certain attribute of the earthquake that took place and the depth can be an indirect reason of an earthquake taking place in that specific location). Given that these locations are irregularly distributed in a geographical space, then this can be thought to be generated by certain stochastic procedure. Hence, the set of locations is a SPP.

Part c

```
ppp_earthquakes<-ppp(earthquakes_italy$Longitude, earthquakes_italy$Latitude, c(min_lon,max_lon), c(min_

# exploratory analysis -- quadrant count
Q <- quadratcount(ppp_earthquakes,
                  nx = 6,
                  ny = 6)


plot(ppp_earthquakes,
     cex = 0.5,
     pch = "+",
     main="Earthquakes locations")
plot(Q,add = TRUE, cex = 2)
```

Earthquakes locations

0	0	0	1	0	1
8	1	2	4	1	1
4	51	4591	1323	9	1
0	6	597	1466	11	0
0	0	0	2	2	0
0	0	0	2	0	2

```
#####################################################
# TESTING FOR CSR (chi-squire test)

ppp_earthquakes.quadtest <- quadrat.test(ppp_earthquakes,
                                         nx = 6,
                                         ny = 6)  # how does number of quadrants affect results?
```

ANSWER

```
ppp_earthquakes.quadtest
```

```
##
##  Chi-squared test of CSR using quadrat counts
##
## data:  ppp_earthquakes
## X2 = 104714, df = 35, p-value < 2.2e-16
## alternative hypothesis: two.sided
##
## Quadrats: 6 by 6 grid of tiles
```

```
print("The actual number of earthquakes in each quadrants are ")
```

```
## [1] "The actual number of earthquakes in each quadrants are "
```

```
ppp_earthquakes.quadtest$observed
```

```
##  [1]    0    0    0    1    0    1    8    1    2    4    1    1    4   51 4591
## [16] 1323    9    1    0    6  597 1466   11    0    0    0    0    2    2    0
## [31]    0    0    0    2    0    2
```

```
print("The expected number of earthquakes in each quadrants are ")
```

```
## [1] "The expected number of earthquakes in each quadrants are "
```

```
ppp_earthquakes.quadtest$expected
```

```
##  [1] 224.6111 224.6111 224.6111 224.6111 224.6111 224.6111 224.6111 224.6111
##  [9] 224.6111 224.6111 224.6111 224.6111 224.6111 224.6111 224.6111 224.6111
## [17] 224.6111 224.6111 224.6111 224.6111 224.6111 224.6111 224.6111 224.6111
## [25] 224.6111 224.6111 224.6111 224.6111 224.6111 224.6111 224.6111 224.6111
## [33] 224.6111 224.6111 224.6111 224.6111
```

There is direct evidence of clustering. We can notice in the quadrant plot that the number of earthquakes is concentrated at the middle of Italy. Now, considering an homogeneous Poisson process would not be appropiate. We notice that the intensity is varying from quadrant to quadrant. The number of expected number of events is 224.6111 if the process were homogeneous Poisson, while the actual number of events completely vary from quadrant to quadrant as shown in the quadrant plot and the values printed out. Furthermore, we can make a chi-squared test, where H0 is that the process is homogeneous process and Ha any other process. When doing the test, we obtain that chi squared statistic value is X2 = 104714 the pvalue p=2.2e-16. Given that the p-value is so small we reject the H0 and accept the Ha. Hence, the earthquake point process in Italy is not Poisson homogeneous. The small p valueal equivalentely means that the stochastic process of the locations of earthquakes is most likely to be of any other kind of stochastic process