

Segunda entrega proyecto probabilidad 2021-I

Germán David Plazas Cayachoa, David Alfonso Oviedo Salamanca, Santiago Rodríguez Morales.

Escuela de Ingeniería, Ciencia y Tecnología, Universidad del Rosario

1. Descripción inicial del problema considerado

Desde que se inició la pandemia por el COVID-19 se ha visto la necesidad de conocer el virus en todas sus facetas. Sabemos por hecho que Colombia es un país muy desigual en el ámbito socioeconómico y sanitario, por lo cual es importante investigar que tanto ha afectado el virus a la salud de los bogotanos teniendo en cuenta diversos factores sociales e individuales de cada persona.

Para el proyecto se utiliza la base de datos extraída de SaludData Bogotá, actualizada al 12 de abril de 2021, siendo una fuente confiable la cual contiene los casos confirmados de COVID-19 en la ciudad [1]. Esta base de datos contiene distintas variables que servirán a la hora de construir un análisis objetivo.

Nuestro problema principal para este proyecto trata de evidenciar una relación entre factores sociales e intrínsecos y el número de casos/fallecidos por COVID-19 en la ciudad de Bogotá. Como un factor influyente social se busca estudiar la relación entre la localidad de residencia y los casos de COVID. Como factores intrínsecos buscamos evidenciar como afecta el COVID a las personas por edad y sexo. Adicionalmente vamos a utilizar un modelo de regresión lineal con el que predeciremos el número de casos acumulados a futuro.

2. Conjunto definitivo de datos seleccionado

Los datos que deseamos analizar se encuentran recopilados en el repositorio GitHub <https://github.com/DavPlazas/ProyectoProbabilidad1> Estos datos serán utilizados para realizar el análisis descriptivo el cual permite obtener representaciones comprensibles que actúan de manera conclusiva a nuestro objetivo principal. En la carpeta Archivos.xlsx se encuentran 7 archivos de Excel en los cuales están distribuidos distintos datos a conveniencia de los análisis descriptivos que serán interpretados.

El archivo Casos.sexo.xlsx contiene el sexo masculino o femenino de todos los pacientes contagiados por COVID.

El archivo `Casos_acumulados_por_fecha.xlsx` contiene las columnas: fecha de diagnóstico, casos por día, casos acumulados y los días transcurridos desde el primer caso detectado. Estos datos sirven para obtener el modelo de regresión lineal de los casos acumulados.

El archivo `ProyectoEdadesMuertos.xlsx` contiene dos columnas con las edades de todos los hombres y mujeres fallecidos respectivamente.

El archivo `TimeD.xlsx` contiene las columnas de nombre de localidad, su respectivo id y el tiempo promedio de diagnóstico en días.

El archivo `promedio_edad_muertos_localidades.xlsx` contiene los datos de la localidad, su id, y el promedio de edad de los muertos por localidad.

El archivo `Numero_casos_x_localidad.csv` contiene los datos de localidad, id, y el número de casos y muertes en cada una de ellas. Estos sirven para obtener el mapa de calor de Bogotá con el número de casos y muertes por localidad.

`Muertes_sexo.xlsx` contiene el sexo masculino o femenino de todos los pacientes fallecidos por COVID.

En la carpeta `archivos` varios se encuentran las imágenes de las gráficas a utilizar y la imagen (polígono) para el mapa de el distrito de Bogotá y sus localidades[2], y en la carpeta `Code` se incluye todo el código en R para el análisis descriptivo.

3. Análisis estadístico descriptivo

En primer lugar, queremos visualizar de una manera amigable el estado de el número total de casos y defunciones por localidad. Para esto, utilizamos el polígono que nos representa el mapa del distrito de Bogotá y sus localidades. Finalmente, generamos un mapa de calor el cual sirve de ayuda para entender mejor el contexto general en el que se encuentra la ciudad dividida por localidades.

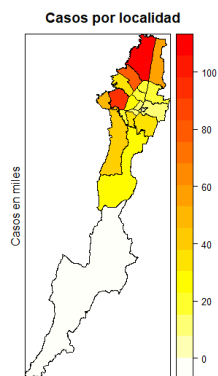


Figura 1: Mapa de calor de los casos positivos por localidad.

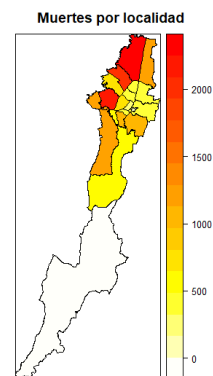


Figura 2: Mapa de calor de los fallecimientos por localidad.

Como segundo punto, queremos analizar que tanto afecta el virus en base al sexo de la persona y ver que tanta diferencia hay entre ambos. Para esto, utilizamos los archivos Casos_sexo.xlsx y Muertos_sexo.xlsx que nos brindan los datos para realizar los diagramas de Torta.

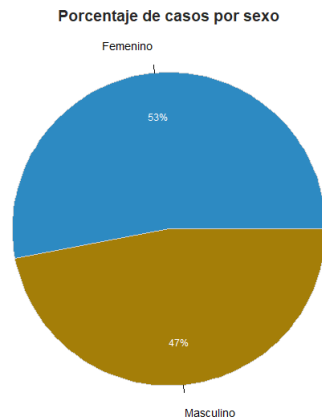


Figura 3: Proporción de casos en hombres y mujeres.

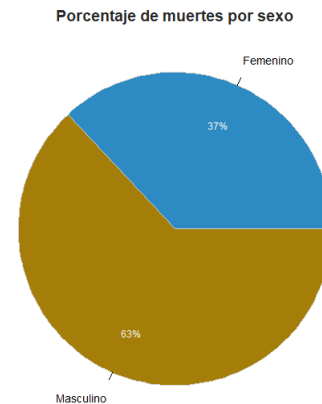


Figura 4: Proporción de fallecimientos en hombres y mujeres.

De esta manera, también queremos ver como se comporta el virus en base a la edad de ambos sexos; visualizando las muertes por grupos de edad y discriminando por hombres y mujeres. Esto lo logramos generando un histograma de dos variables cuyos datos obtuvimos del archivo ProyectoEdadesMuertos.xlsx

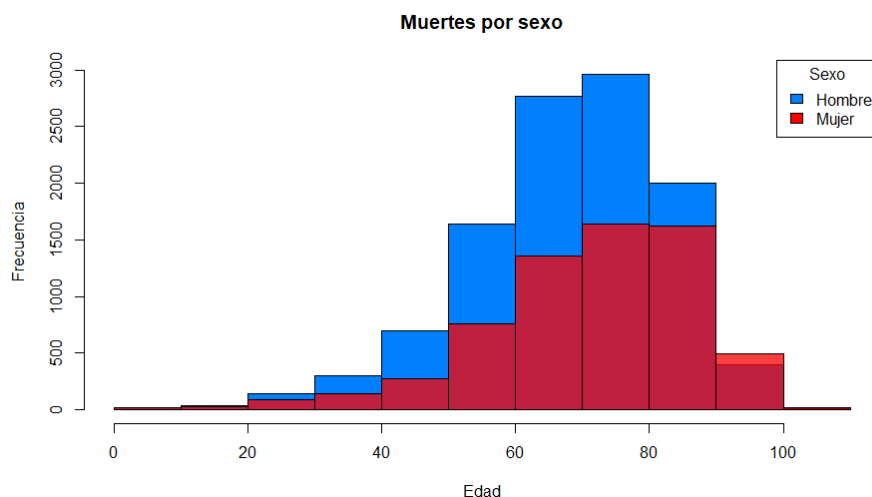


Figura 5: Histograma de muertes por edades y sexos

Para tener una mejor idea sobre los muertos por localidad realizamos dos diagramas de

barras que nos brindan información sobre el número total de defunciones por localidad y también el promedio de edad de los muertos por cada una de las localidades. Con el promedio de edad de cada uno de los fallecidos se busca obtener con mayor precisión información sobre la realidad de diferentes sectores de la ciudad. Para los siguientes diagramas de barras utilizamos los datos de los archivos Numero_casos_x_localidad.csv y promedio_edad_muertos_localidades.xlsx

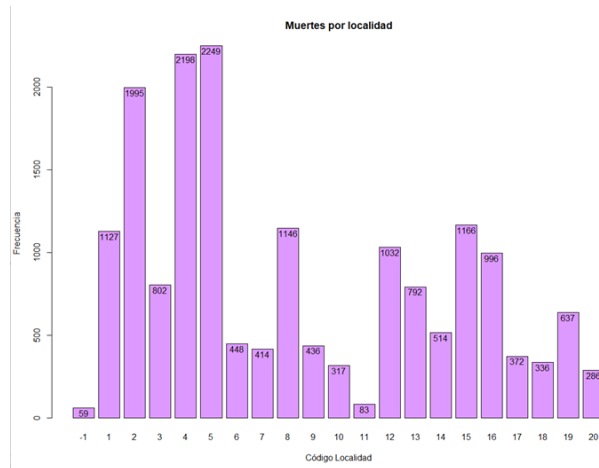


Figura 6: Diagramas de barras número de fallecidos por localidad

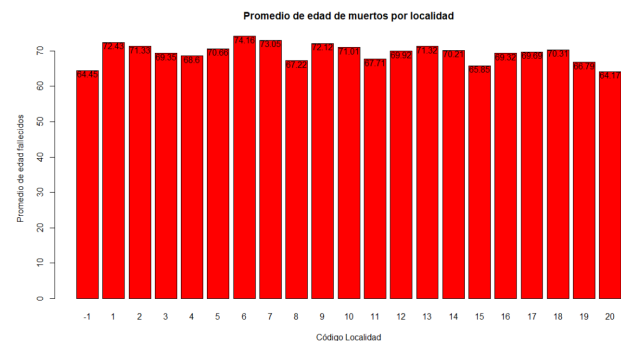


Figura 7: Diagrama de barras promedio de edad de los fallecidos por localidad

- | | | |
|-----------------|-------------------------|----------------------|
| 1: Usaquén. | 8: Cuidad Bolívar. | 15: Bosa. |
| 2: Engativá. | 9: Barrios Unidos. | 16: San Cristóbal. |
| 3: Fontibón. | 10: Los Mártires. | 17: Santa Fe. |
| 4: Kennedy. | 11: La candelaria. | 18: Antonio Nariño |
| 5: Suba. | 12: Rafael Uribe Uribe. | 19: Usme. |
| 6: Teusaquillo. | 13: Puente Arándano. | 20: Fuera de Bogotá. |
| 7: Chapinero. | 14: Tunjuelito. | 21: Sumapaz. |
- 1: Sin dato

Adicionalmente a los análisis descriptivos anteriormente descritos, se decide obtener información que nos brinde una idea más amplia en cuanto a como se ha manejado la detección del virus por cada una de las diferentes localidades de la ciudad, ya que esta puede también variar. Mencionado esto, se decide hacer un diagrama de barras que visibilice el tiempo de diagnóstico promedio en cada una de las localidades ya que esta es una manera sencilla de entender la información y comparar los resultados de cada localidad. Este diagrama lo realizamos utilizando los datos del archivo TimeD.xlsx.

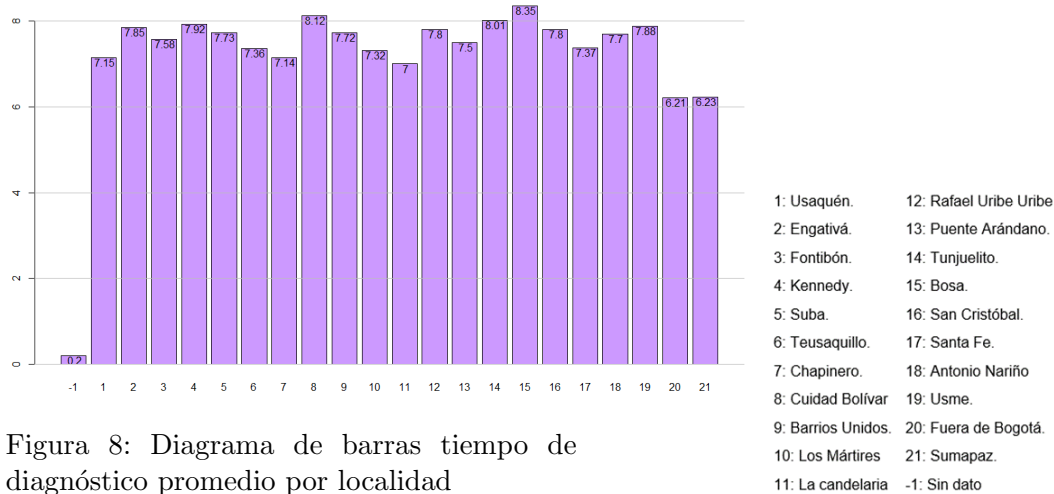


Figura 8: Diagrama de barras tiempo de diagnóstico promedio por localidad

Finalmente, y de manera práctica, resulta conveniente aplicar lo aprendido durante el curso en cuanto al tema de regresión lineal para obtener una predicción de la cantidad de casos acumulados a futuro. Para esto es necesario trabajar con varios datos del archivo Caso-sacumulados_por_fecha.xlsx. Esta información que queremos obtener la vamos a visualizar generando un diagrama de dispersión a lo largo del tiempo (días acumulados) y el total de casos que se van acumulando. Dicho diagrama resulta muy útil para poder entender el comportamiento del virus en la ciudad a lo largo del tiempo.

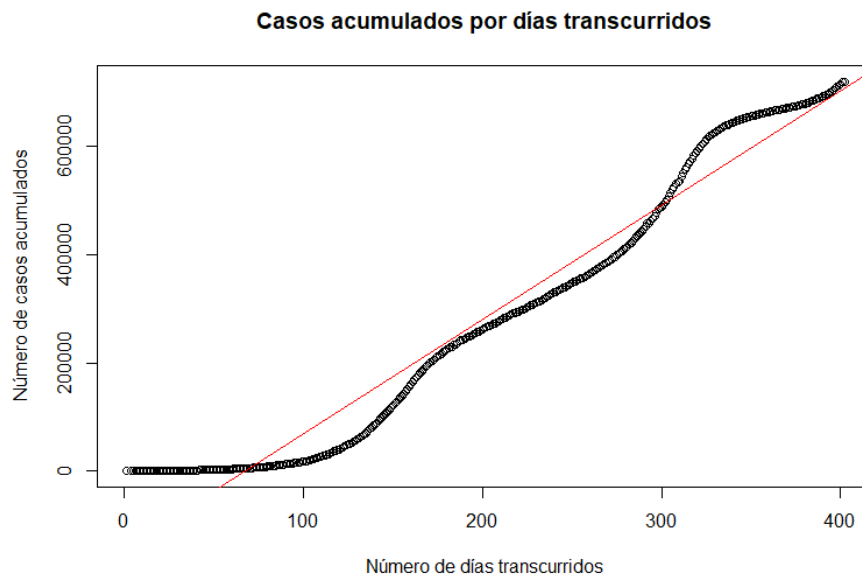


Figura 9: Diagrama de dispersión de regresión lineal. Casos acumulados por días transcurridos

Referencias

- [1]Secretaría de Salud de Bogotá, Casos confirmados de COVID-19 en Bogotá D.C.”, Bogotá, 12 de Abril 2021. Base de Datos. Recuperado de: <https://saludata.saludcapital.gov.co/osb/index.php/datos-de-salud/enfermedades-trasmisibles/covid19/>
- [2]M. Suárez, ”Localidades de Bogotá D.C. - Archivo Carpeta Zip”, Bogotá D.C., 2009. Recuperado de: <https://sites.google.com/site/seriescol/shapes>