

## Primera 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 del Problema

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 es importante saber que tanto ha afectado la pandemia a los bogotanos teniendo en cuenta estas problemáticas. Un factor importante que se desea evidenciar en el proyecto es ver que tanto repercute la edad, el sexo y la localidad en los fallecidos por COVID-19 en la ciudad de Bogotá. Debido a esto, se trabajarán tres hipótesis.

Como primera hipótesis, se plantea que los hombres son más propensos a fallecer por el virus a una edad más temprana que las mujeres. Además, se desea evidenciar que en la localidad de Kennedy la diferencia de mortalidad entre sexos es aún mayor que en Usaquén.

Como segunda hipótesis, se plantea que las personas que viven en las localidades de Usaquén, Chapinero y Suba son más propensas a fallecer a una edad más tardía que las personas de las localidades de Kennedy, Fontibón y Tunjuelito al ser contagiados por el virus.

Como tercera hipótesis, se plantea que el tiempo de diagnóstico desde la aparición de los primeros síntomas es menor en los pacientes de las localidades de Usaquén, Chapinero, Suba y Engativá que en los pacientes de las localidades de Kennedy, Fontibón, Tunjuelito y Ciudad Bolívar.

Para buscar la veracidad de nuestras hipótesis se utiliza la base de datos extraída de SaludData Bogotá, actualizada al 12 de Abril de 2021, que contiene los casos confirmados de COVID-19 en la ciudad [1]. Esta base de datos contiene las variables: Número de caso, fecha de inicio de síntomas, fecha de diagnóstico, localidad, edad, sexo, fuente del contagio, ubicación y estado del paciente.

Teniendo en cuenta lo anterior, en cada hipótesis se utilizarán los siguientes conjuntos de datos:

**Primera Hipótesis:** Sexo, edad y localidad.

**Segunda Hipótesis:** Localidad y edad.

**Tercera Hipótesis:** Localidad y Tiempo de diagnóstico. El tiempo de diagnóstico se puede hallar gracias a la fecha de inicio de síntomas y la fecha de diagnóstico.

## 2. Descripción de las variables de nuestra fuente de datos.

- Número de caso: Variable cuantitativa. Representa el número de caso confirmado por COVID-19.
- Fecha de inicio de síntomas: Variable cualitativa. Representa la fecha del inicio de síntomas de un paciente.
- Fecha de diagnóstico: Variable cualitativa. Representa la fecha en la que el paciente se confirma como un caso COVID-19.
- Tiempo de Diagnóstico: Variable cuantitativa. Representa el tiempo transcurrido entre la fecha de inicio de síntomas y la fecha de diagnóstico. Se mide en días.
- Localidad: Variable Cualitativa. Representa un sector de la ciudad de Bogotá.
- Edad: Variable Cuantitativa. Representa la edad de un paciente. Se mide en años.
- Sexo: Variable Cualitativa. Representa el sexo del paciente. Es F (Femenino) o M (Masculino).
- Fuente del Contagio: Variable Cualitativa. Indica de donde viene el contagio del paciente (Importado, por persona relacionada o desconocido).
- Ubicación: Variable Cualitativa. Indica en donde se encuentra el paciente (Casa, Hospital, UCI, Fallecido).
- Estado del paciente: Variable Cualitativa. Indica el impacto del virus en el paciente (Leve, Moderado, Grave, Recuperado o Fallecido).

## 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/>