

Análisis de gravedad del COVID -19 entre entorno rural y urbano en México

MA1001B – Modelación estadística para la toma de decisiones

Axel Quiroga Caldera A00832676, Sarah Dorado A01540946, Querubín Pérez Andrés
A01761022, Juan Francisco Leal Pérez de Ayala A01285829

22 de octubre del 2022

Abstract

Este análisis aborda el COVID-19 analizando su efecto en mujeres de 35 a 50 años en entorno rural y urbano en México, para hacer el análisis se utilizaron varias técnicas estadísticas, como: prueba de hipótesis para la diferencia de medias, prueba de hipótesis para la diferencia de proporciones y la prueba chi cuadrada para identificar independencia. Se encontró que había diferencia entre las medias de los días hospitalizados en un estado urbanizado y en un estado rural, en el cual el estado rural tenía una media mayor, posteriormente se encontró que el estado rural tenía una mayor proporción de indígenas que el estado urbano, a lo que se realizó una prueba de independencia entre ser indígena y estar intubado (tomando en cuenta la intubación como indicador de mayor gravedad), y se encontró que ser indígena y estar grave son independientes.

Introducción

Durante el transcurso de la pandemia por COVID-19, un gran número de estudios estadísticos han sido publicados, con el propósito de analizar los datos de los afectados por el coronavirus originado en la ciudad china de Wuhan. Un dato clave a analizar es la gravedad en los enfermos de COVID-19. Para ello, se tomó como patrón el tiempo de hospitalización. Habría mucho que matizar en este aspecto, pero presuponemos que, a mayor tiempo de hospitalización, mayor es la gravedad de la enfermedad (y viceversa).

Por otro lado, y como también ha quedado de manifiesto en otros estudios, el COVID-19 no afecta de la misma manera a hombres que a mujeres [en países como Colombia, la mortalidad es mayor en hombres que en mujeres (Cortés, 2020)]. Por ello, se decidió centrar el estudio en uno de los dos sexos, a saber: el femenino. Además, se tomó en cuenta un rango de edad específico de mujeres en distintos estados del país.

Se escogió el entorno urbano y rural, ya que, las diferencias de hábitos característicos entre habitantes de zonas rurales y las más desarrolladas nos han llevado a tomar dos estados similares en este aspecto ["la falta de acceso al agua en zonas urbano-marginales y rurales dificulta poner en práctica medidas de higiene básicas como lavarse las manos con frecuencia (Castro, 2020)"].

Metodología

Para iniciar el análisis, realizamos una investigación previa del COVID-19, porque es cierto que la pandemia fue algo que vivimos, pero hay análisis muy interesantes y que no tienen tanta difusión, este paso previo funciono para desarrollar nuestra pregunta de investigación: *Comparación de tiempo de Hospitalización entre mujeres de 35 a 50 años de entorno urbano y rural, ¿Qué entorno tuvo más días de contagio y cuál pudo haber sido una razón?* Después de haber planteado la pregunta de investigación, se buscó la base de datos y se trabajó con los datos abiertos de la Secretaría de Salud de México, una base de datos con aproximadamente 7 millones de datos. Con esto, se utilizó R para obtener una muestra aleatoria de 120, 000 datos aleatorios.

Después, se seleccionaron las variables de interés para tener un primer acercamiento hacia nuestra pregunta de investigación, las variables seleccionadas fueron: sexo, tiempo de hospitalización (que es la diferencia de dos columnas de los datos fecha de defunción y fecha de ingreso, esto puede significar una limitación en nuestro análisis, ya que no es lo mismo salir recuperado, que fallecer), edades y entidad nacional. Posteriormente se realizó una limpieza de los datos para remover valores atípicos extremos, valores sin sentido y valores nulos, para realizar gráficas y visualizar las variables. Tras realizar el grafico, figura 1 se pude visualizar con los datos que se trabajaran, la media de los días de hospitalización es de 8 días, la media de la edad es de 39 años, la proporción de sexo es similar y los estados de interés, dos urbanos (Nuevo León y Jalisco) y uno rural (Chiapas) tienen una cantidad de paciente similar, por lo que con esto nuestra pregunta de investigación está justificada y se podrá proseguir con la investigación.

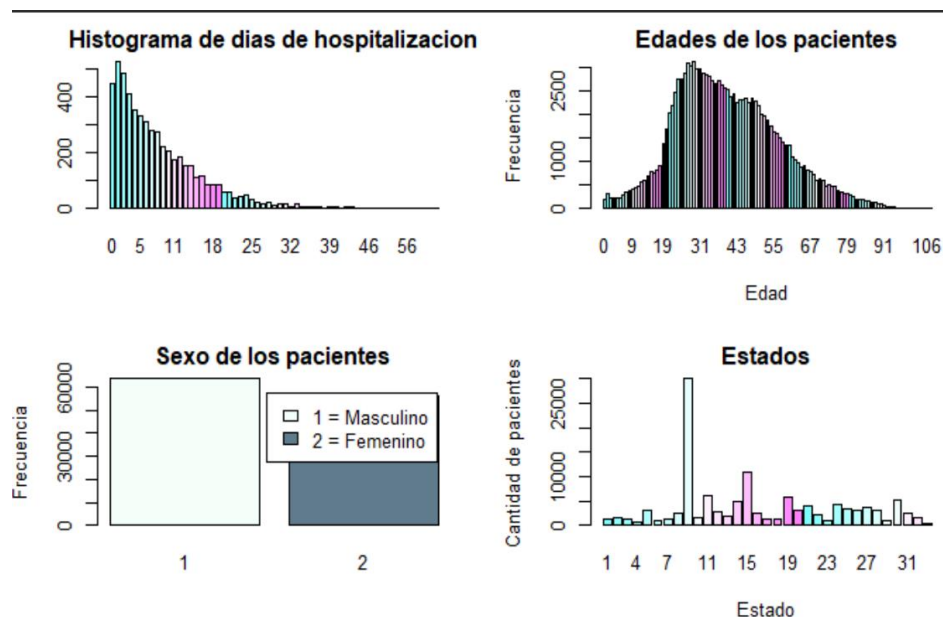


Figura 1. Gráficos de las variables de interés.

Iniciando el análisis inferencial, se tomaron dos estados con un nivel similar de urbanización; Nuevo León y Jalisco, se limpiaron los datos para eliminar valores nulos y seleccionar los datos que contenían a mujeres de 35 a 50 años y resultaron dos muestras con más de 700 datos. Se realizó una diferencia de días de hospitalización para los dos estados y se realizó la primera prueba estadística para obtener la diferencia de medias sin conocer la varianza poblacional. Se plantearon dos hipótesis; *Ho: Las medias son iguales para ambos estados* y *H1: Las medias son diferentes*. Con una significación $\alpha=0.05$, para realizar la prueba de medias se utilizó un z-test utilizando R, ya que son muestras grandes y con varianza desconocida, con eso obtuvimos un intervalo de [-4.52,1.93]; por lo que, se concluyó, al 95% de confianza, que las medias no son diferentes.

Tras saber que las medias no son diferentes, se continuará el análisis inferencial obteniendo las proporciones de indígenas de los dos estados que estamos analizando, para analizar las proporciones se plantearon las siguientes hipótesis; *Ho: Es la misma proporción* y *H1: Es diferente proporción*. Con una significación $\alpha=0.05$, para realizar la prueba se utilizó prop.test en R. Con esto se obtuvo [-0.006 , 0.001], con una confianza del 95% y sabiendo que cero está dentro del intervalo concluimos que la proporción de indígenas es la misma en ambos estados.

Con estos análisis inferenciales, se pudo observar el comportamiento del COVID-19 y su gravedad en estados urbanos en mujeres de 35 a 50 años, pero lo siguiente es realizar los mismos análisis inferenciales para comparar con un estado rural y saber la agresividad del COVID-19.

Para la comparación de un estado urbano y uno rural, se escogió Nuevo León y Chiapas. Se conservaron los datos ya limpiados para el estado de Nuevo León y para Chiapas, se siguió el mismo procedimiento para realizar la limpieza, remover valores nulos y seleccionar a las mujeres de entre 35 a 50 años y la muestra resultante fue de más de 200 observaciones.

Teniendo las muestras de los dos estados, se realizó la prueba de medias sin conocer la varianza poblacional, para esto se plantearon dos hipótesis; *Ho: Las medias son iguales para ambos estados* y *H1: Las medias son diferentes*. Con una significación $\alpha=0.05$, para realizarla se utilizó la función z.test en R, se obtuvo [-6.08,-0.2], no hay evidencia para afirmar que las medias sean iguales, con este resultado también ese resultado indica que la media de días de hospitalización es mayor en Chiapas.

Posteriormente se realizó la prueba de proporción, para saber la proporción de indígenas en ambos estados, se plantearon dos hipótesis; *Ho: Es la misma proporción* y *H1: Es diferente proporción*. Con una significación $\alpha=0.05$, para la prueba se volvió a utilizar prop.test en R. Se obtuvo [-0.499, -0.005], concluyendo que las proporciones de indígenas son diferentes en ambos estados y que Chiapas tiene mayor proporción de indígenas.

Discusión

Tras realizar la comparación de medias de días de hospitalización y de proporciones de habitantes indígenas en los tres estados, se observó que un estado rural tiene más días de hospitalización y más proporción de indígenas, con esto surge la pregunta ¿es cierto que el tiempo de hospitalización dependa, realmente, del hecho de ser indígena? Y para responder esa pregunta, se tomará la intubación como medida de gravedad en la enfermedad.

Posteriormente, se realizó una prueba de independencia, planteando dos hipótesis; *H₀: Ser indígena y estar intubado son independientes* y *H₁: Ser indígena y estar intubado no son independientes*. Con una significación $\alpha=0.05$, posteriormente se utilizó la prueba χ^2 para obtener el p-value. Se obtuvo un p-value= 0.3751 > $\alpha = 0.05$. En consecuencia, concluimos que el hecho de ser intubado no está condicionado por ser o no indígena.

Conclusiones

Tras realizar los análisis pertinentes, se pudo contestar la pregunta de investigación que abordaba una comparación entre la gravedad en entornos diferentes y aunque se encontró que ser indígena y haberse enfermado de gravedad del COVID-19, eso no incluye otros factores que bien si pudieron afectar la gravedad, como las medidas de higiene tomadas en los diferentes estados y como pudo afectar vivir en un entorno rural la gravedad del COVID-19.

Anexo

En la carpeta se encuentra la muestra con la que se trabajó y el markdown con el análisis.

<https://drive.google.com/file/d/1dbL9oAEHZUVvSITvum-ul85SkblhyCtd/view?usp=sharing>

Referencias

Castro, A. (2020, 15 abril). Respuesta a la pandemia de COVID-19 en poblaciones urbano-marginales y rurales en América Latina | Programa De Las Naciones Unidas Para El Desarrollo. UNDP. Recuperado 22 de octubre de 2022, de <https://www.undp.org/es/latin-america/blog/respuesta-la-pandemia-de-covid-19-en-poblaciones-urbano-marginales-y-rurales-en-am%C3%A9rica-latina>

Cortés, A. E. (2020, 1 noviembre). Comportamiento por sexo y género de la pandemia de COVID-19 en Colombia | Revista de Salud Pública. Recuperado 22 de octubre de 2022, de <https://revistas.unal.edu.co/index.php/revsaludpublica/article/view/88913>