



TECNOLÓGICO DE MONTERREY®

Evidencia 1. Análisis estadístico de la base de datos

Análisis de datos para la toma de decisiones

Grupo: 500

Alumnos:

| | |
|------------------------------|-----------|
| Luis Fernando García Mejía | A01367050 |
| Ana Sofia Fuentes Escudero | A01197188 |
| Eduardo Gálvez Vizcaino | A00827197 |
| Luis Eduardo González Quiroz | A01751188 |
| Pedro Alan González Arámbula | A01625308 |

Fecha de entrega: 8 de septiembre del 2020

Contexto

A inicios del 2020, el mundo fue afectado por el esparcimiento del virus COVID-19, en México hubo un cese de actividades presenciales, pero aun así el contagio fue muy extenso, con un incremento exponencial de casos y una tasa de mortalidad de más del 10%.

Conforme avanza la enfermedad se han descubierto patrones y detalles sobre el comportamiento de esta misma. Entre estos descubrimientos se ha visto que existen personas que se encuentran en mayor riesgo de desarrollar sintomatología grave o incluso morir. Dentro de este grupo de personas vulnerables se encuentran quienes tengan ciertas enfermedades subyacentes (cáncer, EPOC, Diabetes Mellitus tipo 2, Hipertensión, Obesidad, entre otras.), adultos mayores y embarazadas. Esto es signo de alerta para México debido a que las principales causas de mortalidad en el país, según las Federación Mexicana de Diabetes, están justamente relacionadas a estos factores de riesgo, donde 19.9% están relacionadas a enfermedades cardíacas y 15.4% a diabetes las cuales muchas veces van a la par de la obesidad (Federación Mexicana de Diabetes, 2018).

El propósito de este análisis es mostrar cual es la relación entre dichos padecimientos, con el número de casos y defunciones por COVID-19 en el país, para buscar acciones a tomar para evitar un peligro mayor a la población mexicana, pero más que eso, nuestro propósito es utilizar dicha información para presentarla en una forma que facilite la comprensión profunda de la misma a la vez que tanto nosotros, como nuestra audiencia, pueda generar conclusiones de alto valor basadas en un análisis objetivo de los datos, contribuyendo a una toma de decisiones informada para navegar de la mejor forma la crisis actual.

Audiencia

Con el desarrollo de este proyecto, y como ya mencionamos antes, nuestro objetivo es poder desarrollar visualizaciones que ayuden a entender la situación actual resultado de la pandemia, e idealmente así poder describir posibles escenarios futuros con el objetivo de tomar mejores decisiones para beneficio de la sociedad. Dicho lo anterior, planteamos que nuestra audiencia sea tan amplia como sea posible para así poder llevar nuestro análisis a la mayor cantidad de personas posibles. En términos prácticos, sin embargo, intentar abarcar una audiencia demasiado amplia es complicado y se corre el riesgo de perder agudeza en el análisis. Por lo anterior decidimos que más en concreto, las organizaciones gubernamentales, médicas y civiles, ONG 's y empresas sociales, serían nuestro público objetivo. Así, podemos ahondar en nuestro análisis, mismo que dichas organizaciones podrán entender y aprovechar a mayor profundidad, para que así éstas a su vez, deriven los beneficios de nuestro análisis a la sociedad en general.

Problemática

La problemática que intentamos resolver con este desarrollo es que, si bien hay mucha información respecto a la situación actual con respecto a la pandemia, dicha información se presenta y se comparte con métodos y paradigmas que hacen difícil que organizaciones, particulares y la sociedad en general puede tener acceso a verdadero conocimiento que pueda ser útil para una toma de decisiones informada. Notamos que hay mucha información, pero poco conocimiento que agregue valor a dicha información. Es por ello, que nuestro propósito se centra en pre procesar esa información, y presentarla primero, en un formato que facilite la comprensión profunda de la misma a la vez que incentiva la generación de conclusiones útiles para la sociedad al momento de lidiar con la crisis actual. Nuestra hipótesis inicial sobre esta problemática es que hay una estrecha relación sobre la fatalidad del COVID con los padecimientos que ya tiene el paciente.

Justificación de elección de variables

Para la realización de este análisis fueron escogidas 4 variables en total, 2 variables cuantitativas y 2 variables cualitativas, la selección de estas fue pensada para poder realizar una relación entre las mismas, específicamente se busca revisar si ciertas condiciones aumentan el riesgo o la probable letalidad del COVID-19. Decidimos realizar esta relación debido a que, al explorar la base de datos, nos dimos cuenta de que un porcentaje bajo de las defunciones eran solo por COVID, es decir que las demás personas que fallecieron tenían una o más situaciones más además del COVID, por lo que elegimos variables que nos mostrarían esta relación.

Análisis estadístico de las variables

Variables cuantitativas

Defunciones por estado

Hasta el 28 de Julio se promediaron 1854 muertes por estado debido al COVID-19, se observó durante este mismo periodo que la mitad de los 32 estados de la república presentó una cifra igual o menor a 1172 defunciones. Podemos observar de mejor manera la distribución de defunciones por estado con las medidas de ubicación: Observamos que el 75%, es decir, 24

Estadísticas

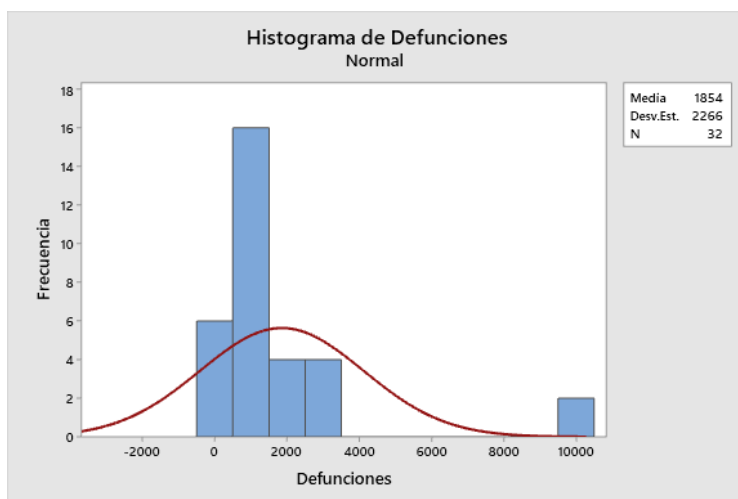
| Variable | Media | Mediana | Modo | N para moda |
|-------------|-------|---------|------|-------------|
| Defunciones | 1854 | 1172 | * | 0 |

Estadísticas

| Variable | Desv.Est. | Varianza | CoefVar |
|-------------|-----------|----------|---------|
| Defunciones | 2266 | 5133864 | 122.20 |

Estadísticas

| Variable | Q1 | Mediana | Q3 |
|-------------|-----|---------|------|
| Defunciones | 605 | 1172 | 2187 |



estados presentaron más de 605 defunciones, siendo un 25% (8) de los estados que superaron las 2187 muertes.

Asimismo, sabemos que el valor promedio de defunciones presenta una variabilidad de menos o más 2266.

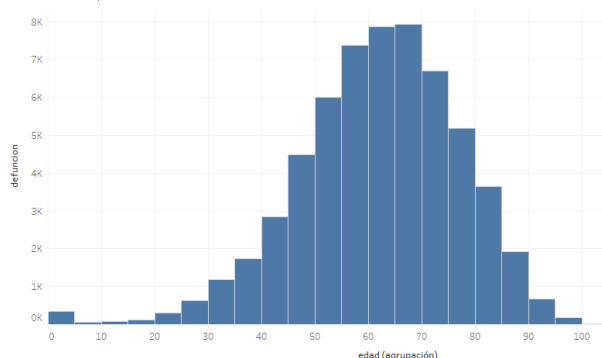
Gracias al histograma de defunciones, sabemos que seis estados presentaron una cifra de muertes entre un mínimo de 237 a 500, mientras 16 estados, la mayoría, presentaron una cifra de entre 500 y 1500 defunciones, cuatro estados una cantidad de entre 1500 y 2500, una misma cantidad de estados presentaron en 2500 y 3500. Por último, solamente dos estados presentaron entre 9500 y 10500.

Edad

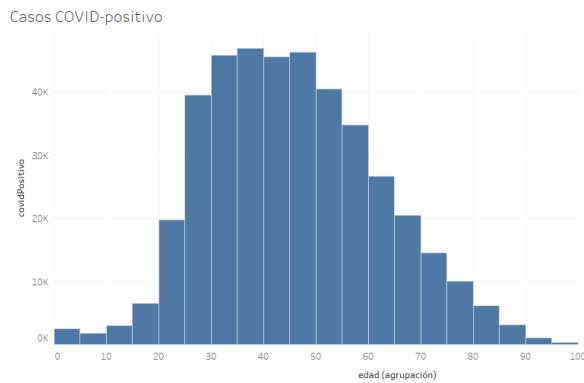
Decidimos utilizar la edad como variable cuantitativa pues

consideramos que era importante para entender la naturaleza del virus y cómo interactúa con las personas. Después de hacer nuestro análisis, descubrimos varios datos particularmente importantes. En primer lugar, si vemos la gráfica de personas que reportaron COVID-positivo, vemos que en términos absolutos hay una marcada tendencia en personas arriba de 30 años a contraer el virus. Si bien este dato ya nos dice mucho, si revisamos una simple pirámide poblacional de México, podemos ver que la mayoría de la población en México está concentrada en grupos de edad jóvenes, por lo que tener en total una cantidad de casos

Defunciones por COVID



positivos tan pequeños en esos grupos es aún más remarcable si lo vemos desde una perspectiva relativa. Después, revisamos el número de defunciones por COVID [3] notamos el mismo fenómeno que ya



mencionamos anteriormente, la mayoría de los casos están concentrados en grupos de edad avanzados, ahora en particular arriba de 45 años, lo que, de nuevo es todavía más remarcable tomando en cuenta las distribuciones de edad en México. En conclusión, analizando la edad de nuestro grupo de estudio, vemos que si bien en general la población de México está

concentrada en grupos jóvenes (entre 0 y 30 años), la distribución de casos positivos de COVID se concentró principalmente en grupos de edad más avanzados (arriba de 30 años), y en el caso de defunciones por COVID, se carga aún más hacia grupos de edad mayor (arriba de 45 años). De esta forma podemos ver que si bien, el virus afecta a todos, los grupos de edad que se encuentran en más alto riesgo de contagio o muerte son aquellos arriba de 40 años.

Variables cualitativas

Neumonía

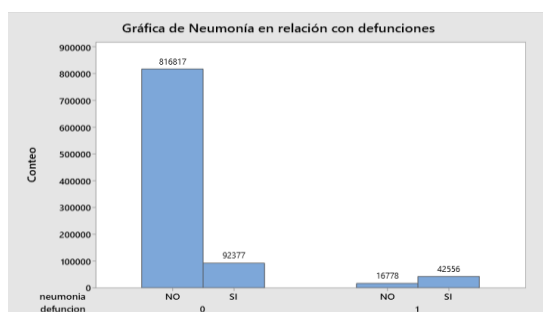
Contar

| neumonía | Conteo | Porcentaje |
|----------|--------|------------|
| NO | 833595 | 86.07 |
| SI | 134933 | 13.93 |
| N= | 968528 | |

La variable de neumonía es aplicable para toda la población de la base de datos, aunque esto no ayuda realmente con el propósito de este análisis, interpretamos la tabla de frecuencias en que un 14% de la población cuenta con esta enfermedad, siendo un gran número de

personas.

Ahora para poder darle valor al análisis se debe relacionar con otra variable y se interpreta, en este caso como se observa en la gráfica de barras se utilizó la variable de defunciones por COVID-19. En esta gráfica de barras tenemos una relación entre la cantidad de defunciones por COVID en las que el paciente tenía neumonía (0 siendo que está vivo y 1 defunción confirmada) Se observa que el número de defunciones es relativamente alto cuando el paciente presenta neumonía, estadísticamente hablando podemos notar que el COVID-19



presenta un riesgo a la vida más alto cuando se junta con neumonía. Esta representación gráfica fue usada ya que es notablemente más fácil observar la relación entre dos variables cuando se usan las gráficas de barras, esto porque se cuenta con el conteo dividido y fácil de entender.

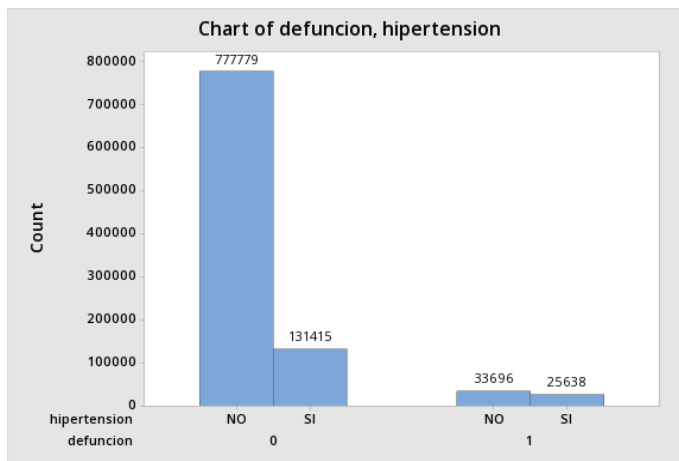
Hipertensión

Tally

| hipertension | Count | Percent |
|--------------|--------|---------|
| NO | 811475 | 83.78 |
| SI | 157053 | 16.22 |
| N= 968528 | | |

Al igual que la variable de neumonía, la variable de hipertensión es aplicable para todos los usuarios de la base de datos sin embargo tomamos en consideración para este análisis únicamente a quienes presentan este padecimiento siendo el 16.2% de la población. Este porcentaje representa un alto número de personas con la enfermedad, incluso mayor al número de personas con neumonía. Utilizamos esta variable y la relacionamos con el número de defunciones por COVID-19 (como se muestra en la gráfica de barras presentada a continuación) para darle valor al análisis.

En esta gráfica de barras relacionamos el número de defunciones causadas por COVID-19 en las que el paciente además padecía de hipertensión (siendo 0 pacientes aún con vida y



siendo 1 defunciones confirmadas).

Podemos notar que el número de pacientes que sufren de hipertensión y fallecen es relativamente menor al número de pacientes que la padecen y permanecen con vida, sin embargo, es poca la diferencia por lo que podemos concluir que el padecer de hipertensión incrementa el riesgo de muerte al contraer COVID-19. Este

tipo de gráficas permite una mejor

visualización y entendimiento de los datos, y en este caso fue utilizada ya que permite ver la relación que existe entre dos variables.

Descubrimientos

Pudimos observar con este análisis que nuestra hipótesis inicial se cumplió ya que hay una estrecha relación entre las defunciones y factores que varían desde el grupo de edad al que pertenecía el paciente, el estado en el que habitaba, así como el padecimiento que presentaban los pacientes en el caso de neumonía e hipertensión. La relación se hizo evidente al observar las distintas representaciones estadísticas y gráficas que se presentaron anteriormente.

En primera instancia, descubrimos que la mayor cantidad de muertes se encontraba en la zona centro del país, siendo el Estado de México y la Ciudad de México quienes tuvieron una mayor cifra. A continuación, el análisis de las defunciones por grupo de edad arrojó que, mientras la edad de los pacientes aumentaba, había más probabilidad de morir por la enfermedad, siendo el grupo a partir de los 60 años el más afectado. Finalmente, pudimos apreciar el factor de riesgo que representa contraer COVID cuando se tiene un padecimiento como la neumonía o la hipertensión; En el caso de neumonía, 42,556 de las personas que la padecían murieron, en comparación con la 16,778 que la presentaron y sobrevivieron, un 71.7% y 28.3% respectivamente. Por otra parte, 25,638 de los 59,334, es decir, el 43% de pacientes que padecían de hipertensión fallecieron.

Es de vital importancia identificar las relaciones entre la enfermedad y los distintos factores que llevan al paciente a la muerte, haciendo esto, se pueden anticipar varios resultados. Al día de hoy, gracias a las herramientas estadísticas es que sabemos que la enfermedad tiene más probabilidades de terminar con la vida de adultos, o que ciertos padecimientos hace que se sea más difícil combatir a la enfermedad; La naturaleza de la enfermedad, y el tiempo tan breve que ha tardado en transformar la sociedad nos obliga a describir la enfermedad por medio de la estadística, y si bien, todavía no se da con alguna cura, si se pueden prevenir situaciones que podrían resultar mortales para cierto tipo de la población mexicana.

CONCLUSIÓN INDIVIDUAL.

Tras haber realizado todo el análisis estadístico de las variables y determinar que sí existen una relación de las defunciones con las otras variables nos podemos dar cuenta la importancia de las correlaciones pues si lo vemos en este caso en particular, podemos determinar los sectores que están más en peligro ante esta situación lo cual nos ayuda tanto para prevenir como para combatir a la enfermedad pues al saber las características del paciente es más fácil poder crear planes para enfrentar este virus y así poder ayudar a todas las personas de acuerdo a sus condiciones, ya sean padecimientos o edades.

Ahora si hablamos en una generalización la estadística toma una importancia no solo en este sector pues esta rama tiene aplicaciones en diversos campos, el encontrar las correlaciones

no ayuda a determinar sucesos, poder prevenirlos, explicarlos, predecirlos, aquel que conoce las variables y como estas provocan cambios en el mundo es aquel capaz de tomar las decisiones y es aquí donde la ciencia de datos brilla en todo su esplendor, pues este relativo nuevo campo de estudio es la que nos permite entender nuestro mundo y tomar decisiones en base a los millones y millones de datos generados al día alrededor del mismo, haciendo más sencillo todo el proceso de toma de decisiones es extremadamente impresionante como el mundo y sus miles de variables son capaces de describir de forma cuantitativa y cualitativa el como funciona el mundo y como debemos afrontarlo.

Bibliografía

Enfermedad del coronavirus 2019 (COVID-19). (2020). Centers for Disease Control and Prevention.
https://espanol.cdc.gov/coronavirus/2019-ncov/need-extra-precautions/people-with-medical-conditions.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fneed-extra-precautions%2Fgroups-at-higher-risk.html
Federación Mexicana de Diabetes AC. (2018). Estadísticas en México - Federación Mexicana de Diabetes. Fmdiabetes.Org/. <http://fmdiabetes.org/estadisticas-en-mexico>