

Análisis estadístico: el COVID-19 y su relación con la diabetes

Modelación estadística para la toma de decisiones

Isaac Alexis López Paredes - A01747148

Rubén Darías Castro Terrazas - A00833945

César Castro Quintero - A01742286

Jorge Raúl Rocha López - A01740816

Tecnológico de Monterrey

23 de octubre de 2022

Análisis estadístico: el COVID-19 y su relación con la diabetes

Resumen

El presente trabajo analiza el impacto de la diabetes en la tasa de supervivencia de personas con COVID-19 en México de 2020 a 2022 a través del análisis de datos. Se realizó un análisis descriptivo y un análisis inferencial, tanto univariado como multivariado. Se realizaron gráficos, intervalos de confianza y pruebas de hipótesis utilizando R (versión 4.2.1, 2022-06-23). Con un nivel de confianza de 95%, se obtuvieron como resultados principales que durante 2022 las personas sin comorbilidades tienen entre 78% y 79% más de sobrevivir que los que tienen diabetes; esta diferencia ha aumentado desde el inicio de la pandemia y la proporción de sobrevivientes ha aumentado desde 2020 para las personas sin comorbilidades, pero no para las personas con diabetes.

Introducción

A finales del año 2019 surgió una nueva enfermedad en Wuhan, China causada por el virus SARS-CoV-2. De acuerdo con la Organización Mundial de la Salud, “La mayoría de las personas que son infectadas por el virus, presentarán una enfermedad respiratoria leve a moderada y se podrán recuperar sin necesitar un tratamiento especial. Sin embargo, las personas que cuentan con una edad avanzada o alguna comorbilidad son más propensos a desarrollar una enfermedad grave” (OMS, 2020). En México, una de las comorbilidades con más presencia es la diabetes y en 2020, pasó a ser la tercera causa de muerte en el país (INEGI, 2021). Por ello, el presente trabajo tiene como objetivo analizar el impacto que esta condición en específico tiene sobre la supervivencia de un paciente de COVID-19. Adicionalmente se quiere conocer si las vacunas impactaron positivamente tanto a los pacientes con diabetes como a aquellos sin ninguna comorbilidad. Lo anterior se hará a través de un análisis estadístico, guiado por la siguiente pregunta de investigación: De los

clasificados como positivos a COVID, ¿en qué proporción las personas sin comorbilidades tienen más posibilidad de sobrevivir que las que tienen diabetes en lo que va de este año?

¿Es la misma proporción que la de años anteriores? Adicionalmente, ¿los datos dan argumentos para decir que las vacunas tuvieron un impacto en esta proporción?

Metodología

Variables asociadas

Las variables a través de las cuales se abordará la pregunta rectora se dividen en dos categorías. Las primeras estaban presentes desde el inicio en la base de datos:

1. Diabetes. Indica si el paciente cuenta con un diagnóstico de diabetes o no.
2. Defunción. Indica si el paciente sobrevivió o no.

Las siguientes fueron agregadas con el fin de realizar el estudio de manera adecuada:

3. Año. Esta variable se creó a partir del año en que el paciente falleció, o en el caso de que no fuera así, el año en el que comenzaron sus síntomas.
4. Sin comorbilidades. Incluye a todos los pacientes que no tienen ninguna de las comorbilidades presentes en el registro: asma, hipertensión, enfermedades cardiovasculares, EPOC, insuficiencia renal crónica, obesidad, tabaquismo, inmunosupresión u otras comorbilidades.

Descripción de los datos

Los datos que se analizaron son procedentes de la Secretaría de Salud Mexicana. Incluye alrededor de 7 millones de registros de personas con COVID confirmado (al 14 de septiembre de 2022). Se obtuvo una muestra de 120 mil pacientes a través de un muestreo aleatorio estratificado por sexo, para conservar la misma proporción de hombres y mujeres de la población en la muestra.

Para la limpieza de los datos, se quitaron elementos vacíos y se filtraron los datos a partir de varios criterios. Esto con el objetivo de organizar la información de manera que fuera útil para el análisis de las variables asociadas a la pregunta de investigación. Es importante mencionar que el nivel de confianza que se utilizó durante el estudio fue del 0.05.

Análisis Descriptivo de las Variables

Tablas de contingencia

Una vez que se contaban con los datos filtrados, fue posible obtener las primeras observaciones relevantes para la pregunta de investigación. En la Tabla 1 se muestra el número de sobrevivientes con diabetes para cada año. Esta tabla fue utilizada posteriormente para realizar una prueba de independencia.

Tabla 1.

Sobrevivientes con diabetes

	2020	2021	2022
Sí sobrevivió	2633	2633	2808
No sobrevivió	999	938	160

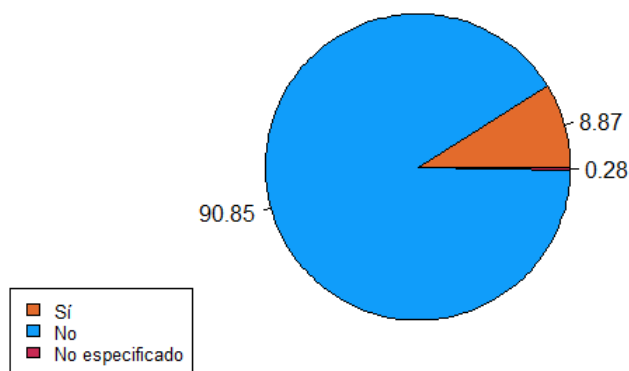
Por otro lado, la Tabla 2 es una tabla cruzada de proporciones. Compara los sobrevivientes con diabetes y sin comorbilidades en cada año. Desde este punto del análisis parece razonable pensar que es menos probable que las personas con diabetes sobrevivan, pues se observa una gran diferencia entre la tasa de supervivencia de la muestra

Tabla 2.*Proporción de sobrevivientes*

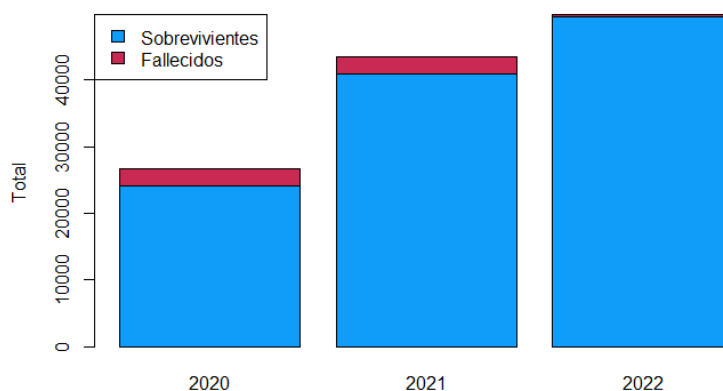
	Con diabetes	Sin comorbilidades
2020	0.10960786	0.74631588
2021	0.06434192	0.82293143
2022	0.05682370	0.85000809

Gráficas

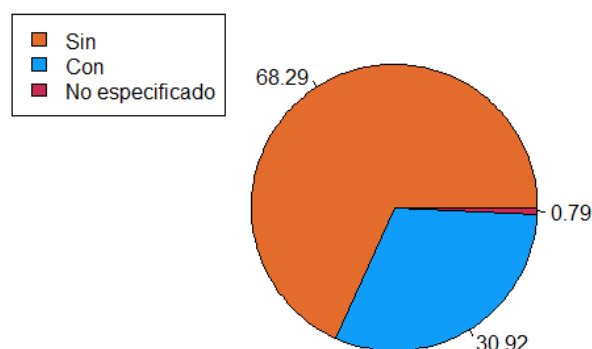
A continuación, se presenta de manera visual la exploración de las variables asociadas. Primero se presentan los gráficos de una sola variable. En la figura 1 se observa la proporción de pacientes (en %) de la muestra con, sin diabetes y no especificado.

Figura 1.*Proporción de pacientes diagnosticados con diabetes*

La figura 2 es un gráfico de relación entre las variables de defunción y año, en el que se aprecian las diferencias entre los sobrevivientes y fallecidos, así como su comportamiento a través de los años. Este parece indicar que la cantidad de contagiados ha aumentado, la de sobrevivientes también y la de fallecidos disminuido.

Figura 2.*Número total de pacientes por año*

Para la variable de comorbilidades se realizó una gráfica de pastel para observar la proporción de pacientes con alguna comorbilidad y los pacientes sin ninguna comorbilidad. En la figura 3 se observa que casi un tercio de la muestra tiene alguna comorbilidad.

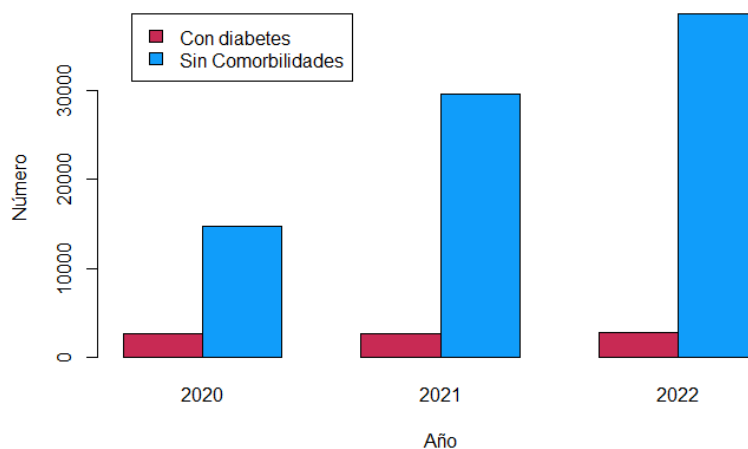
Figura 3.*Proporción de pacientes con y sin comorbilidades*

A continuación, se presentan los gráficos de relación más relevantes, pues incluyen la exploración de todas las variables. En la figura 4 se observa que, en los tres años analizados, la cantidad de personas que no tenían comorbilidades que sobrevivieron fue

mayor que las que tenían diabetes. Además, parece que este número se mantuvo constante para el caso de los diabéticos, pero que creció para las personas sin comorbilidades.

Figura 4.

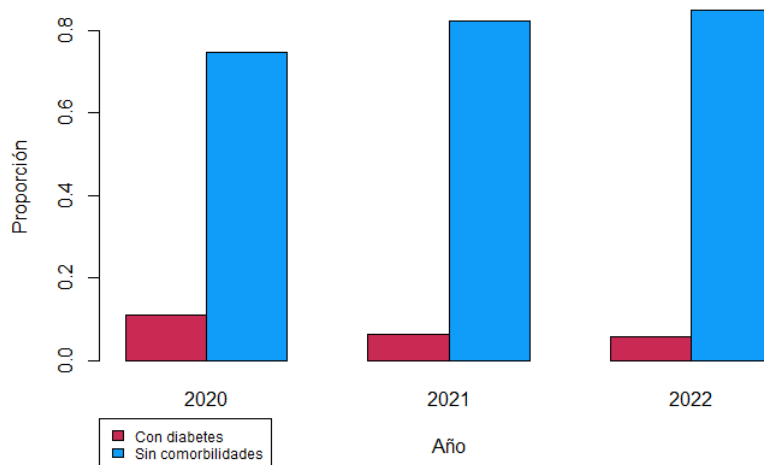
Número de sobrevivientes de 2020 a 2021



En cuanto a las proporciones, parece que estas siguen un comportamiento parecido. En la figura 5 se observa que la proporción de sobrevivientes sin comorbilidades fue mayor que la de las personas con diabetes. Resalta también que parece que la proporción de sobrevivientes con diabetes en 2021 disminuyó con respecto a la de 2020.

Figura 5.

Proporción de sobrevivientes en 2020, 2021 y 2022



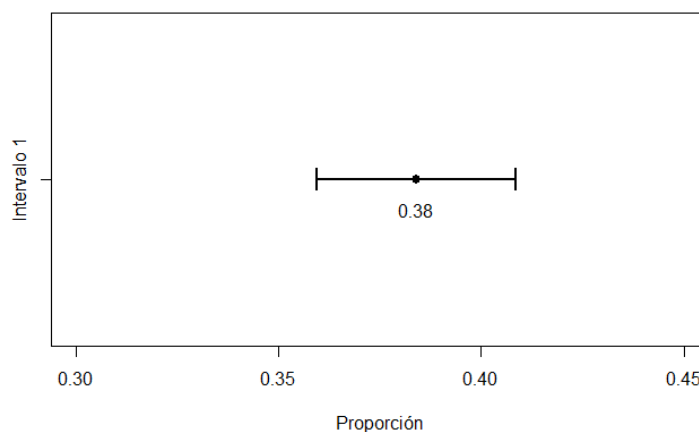
Análisis inferencial

Intervalos de confianza

A continuación, se presentan los intervalos de confianza realizados, relevantes para responder la pregunta de investigación. En primer lugar, se quiere estimar la verdadera proporción de fallecidos con diabetes en 2020. En una noticia publicada por la UDG se encontró que “siete de cada diez muertes de personas relacionadas con el COVID-19 están vinculadas con casos de diabetes” (Loera, 2020), lo cual parece ser una cifra elevada. En la figura 3 se observa la estimación de la proporción poblacional de los pacientes diagnosticados con diabetes y que fallecieron en 2020. La estimación puntual es de 0.38 y por medio de una prueba de proporción se puede inferir que la verdadera proporción se encuentra entre $[0.359, 0.408]$.

Figura 6.

Proporción de fallecidos con diabetes en 2020

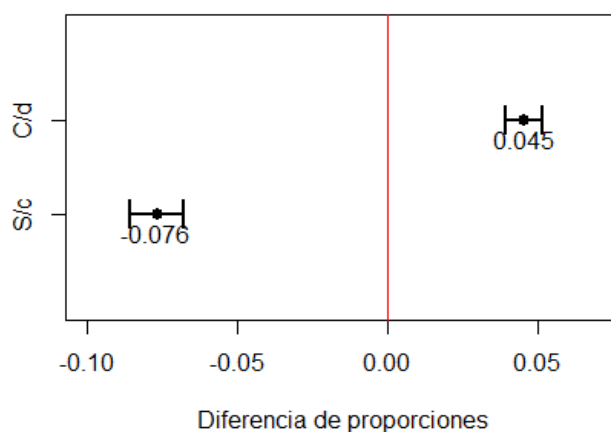


De especial interés es conocer si ocurrió algún cambio en la proporción de sobrevivientes a partir de 2021, pues fue el año en que comenzó la campaña de vacunación. Además, anteriormente se comentó que parece inusual que la proporción de sobrevivientes con diabetes haya disminuido, por lo que se busca comprobar que esto haya sucedido. En la

figura 7 se puede observar que, para el caso de los pacientes con diabetes, muy probablemente la proporción de sobrevivientes fue mayor en 2020, pues la diferencia se encuentra en un intervalo positivo. Por otro lado, para los pacientes sin comorbilidades, lo más probable es que en 2021 la proporción de sobrevivientes haya aumentado en comparación con el 2020, ya que la diferencia se encuentra en un intervalo negativo.

Figura 7.

Diferencia de proporciones de sobrevivientes de 2020 a 2021

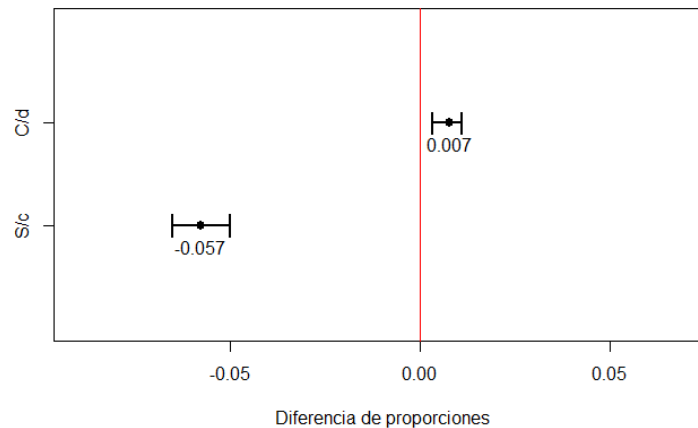


Nota. El intervalo superior muestra la diferencia entre las tasas de 2020 y 2021 para las personas con diabetes y el inferior la diferencia entre las tasas de 2020 y 2021 para las personas sin comorbilidades.

El mismo análisis se realizó para la diferencia entre los años 2021 y 2022. En la figura 8 se observa que sucedió el mismo fenómeno que de los años 2020 a 2021, sin embargo, para el caso de las personas con diabetes, la diferencia de proporciones entre ambos años fue muy pequeña (entre el 0.3% y 1%).

Figura 8.

Diferencia de proporciones de sobrevivientes de 2021 a 2022

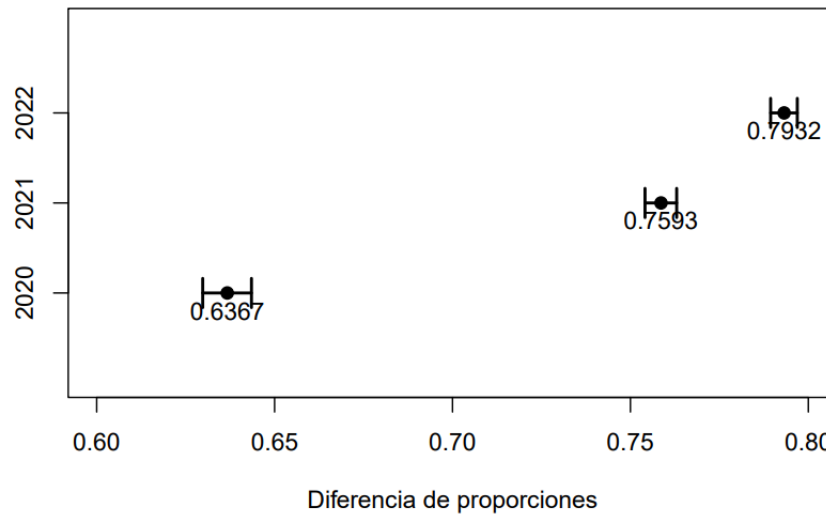


Nota. El intervalo superior muestra la diferencia entre las tasas de 2021 y 2022 para las personas con diabetes y el inferior la diferencia entre las tasas de 2021 y 2022 para las personas sin comorbilidades.

Ahora bien, para comparar las tasas de supervivencia de las personas sin comorbilidades y de las personas con diabetes, se realizaron intervalos para sus diferencias de proporciones. En la figura 9 se observa que muy probablemente la tasa de supervivencia para las personas sin comorbilidades es mayor que para las personas con diabetes. Con el fin de comparar esta diferencia a través de los años, se incluyen tres intervalos en la figura, una para cada año. Así se puede apreciar que esta diferencia ha aumentado con los años. Lo anterior parece razonable puesto que como se comentó en los intervalos anteriores, muy probablemente la tasa de supervivencia para las personas sin comorbilidades ha aumentado, mientras que para la de personas con diabetes ha disminuido (aunque en magnitud pequeña).

Figura 8.

Diferencia de proporciones de sobrevivientes sin comorbilidades y sobrevivientes con diabetes en 2020, 2021 y 2022



Pruebas de hipótesis

Independencia de sobrevivir o no con diabetes

Esta prueba era importante para verificar el supuesto más importante de la investigación, es decir, que la diabetes es un factor relevante para la supervivencia de un paciente de COVID. Para ello se realizó una prueba de independencia utilizando la tabla de contingencia bivariada de los pacientes con diabetes que sobrevivieron o no en toda la muestra (tabla 1). Las hipótesis por probar son las siguientes:

H_0 : Sobrevivir es independiente de diabetes

H_1 : Sobrevivir no es independiente

Y la regla de decisión que se estableció es que se rechaza H_0 si:

$$\text{Regla clásica: } |\chi^2_*| > \chi^2_o$$

Regla del valor p : $p < 0.05$

Con la información de la Tabla 1, se realizó una prueba de independencia y se obtuvieron los siguientes resultados:

$$|\chi_*^2| = 5867 > 3.8414 = \chi_o^2$$

$$p < 2.2 * 10^{-16} < 0.05$$

Por lo tanto, se rechaza H_o y se concluye que tener diabetes no es independiente de sobrevivir con un 95% de confianza.

Proporción de fallecidos en 2020

Esta prueba complementa lo encontrado con el intervalo de confianza presentado en la figura 6. Las hipótesis por probar son las siguientes:

$$H_o: p = 0.7$$

$$H_1: p < 0.7$$

Se obtuvo un valor de p muy por debajo del 0, por lo que se rechaza la hipótesis nula y se concluye que muy probablemente la verdadera proporción sea menor a la que afirma la noticia citada, lo cual confirma lo encontrado en el intervalo de confianza.

Diferencia de proporciones

En la tabla 3 se resumen las pruebas de hipótesis realizadas para diferencia de proporciones. Para todos los casos se utilizó una prueba de igualdad de proporciones con el estadístico X^2 . Se puede apreciar que en todos los casos, se rechazó la hipótesis nula. En la

siguiente sección se interpretan estos resultados para dar respuesta a la pregunta de investigación.

Tabla 3.

Resultados para las pruebas de hipótesis de diferencia de proporciones

Proporción 1	Proporción 2	Nula	Alternativa	Valor p	Decisión
Sobrevivientes con diabetes en 2020	Sobrevivientes con diabetes en 2021	$p_1 - p_2 = 0$	$p_1 - p_2 > 0$	< 0.05	Rechazar la nula
Sobrevivientes sin comorbilidades en 2020	Sobrevivientes sin comorbilidades en 2021	$p_1 - p_2 = 0$	$p_1 - p_2 < 0$	< 0.05	Rechazar la nula
Sobrevivientes con diabetes en 2021	Sobrevivientes con diabetes en 2022	$p_1 - p_2 = 0$	$p_1 - p_2 > 0$	< 0.05	Rechazar la nula
Sobrevivientes sin comorbilidades en 2021	Sobrevivientes sin comorbilidades en 2022	$p_1 - p_2 = 0$	$p_1 - p_2 < 0$	< 0.05	Rechazar la nula
Sobrevivientes en 2020 sin comorbilidades	Sobrevivientes en 2020 con diabetes	$p_1 - p_2 = 0$	$p_1 - p_2 > 0$	< 0.05	Rechazar la nula
Sobrevivientes en 2021 sin comorbilidades	Sobrevivientes en 2021 con diabetes	$p_1 - p_2 = 0$	$p_1 - p_2 > 0$	< 0.05	Rechazar la nula
Sobrevivientes en 2022 sin comorbilidades	Sobrevivientes en 2022 con diabetes	$p_1 - p_2 = 0$	$p_1 - p_2 > 0$	< 0.05	Rechazar la nula

Discusión y Conclusiones

A partir del análisis se puede decir que se cuenta con suficiente evidencia estadística, con un nivel de confianza del 95%, para concluir lo siguiente:

1. Sobrevivir o no, es dependiente de tener diabetes
2. En 2020, entre el 35% y 40% de las personas que fallecieron tenían diabetes, que, aunque es mucho menor a la de 70% que se encontró en la noticia citada, sigue siendo de magnitud importante.
3. En lo que va de 2022, las personas sin comorbilidades tienen entre 78% y 79% más de sobrevivir que los que tienen diabetes.
4. Esta diferencia de proporciones ha aumentado significativamente desde el inicio de la pandemia, pues era entre 62% y 64% en 2020 y entre 75% y 76% en 2021.
5. A pesar de que en 2021 se comenzaron a aplicar vacunas, los datos indican que estas no tuvieron el mismo efecto positivo para los diabéticos que para las personas sin comorbilidades:
 - a. La proporción de sobrevivientes con diabetes en 2021 fue menor que en 2020 y la de 2022 fue menor que en 2021
 - b. La proporción de sobrevivientes con diabetes en 2021 fue mayor que en 2020 y la de 2022 fue mayor a la de 2021.

Con base en estas conclusiones se puede decir que es de suma importancia tomar medidas para prevenir y tratar la diabetes en nuestro país. Aún a pesar de las vacunas, tener diabetes es un importante factor de riesgo para las personas con un diagnóstico de COVID. El presente análisis da pie a seguir investigando la relación entre las variables mencionadas. En particular sería importante considerar los meses en que se aplicaron las vacunas en cada estado de la República, con el fin de analizar de mejor manera la utilidad de éstas para las

personas con diabetes. Además, a partir de este trabajo surgen nuevas preguntas de investigación relevantes. Por ejemplo: ¿cuál es el efecto de tener diabetes sobre los días de hospitalización por COVID?, o ¿cómo se compara la probabilidad de supervivencia estimada con la de pacientes con otras comorbilidades?

Referencias

- INEGI. (12 de Noviembre de 2021). *Estadísticas a propósito del día mundial de la diabetes*. Obtenido de https://www.inegi.org.mx/contenidos/saladeprensa/aproposito/2021/EAP_Diabetes2021.pdf
- Loera, M. (3 de Noviembre de 2020). *En México, la mayoría de fallecidos por Covid tenían diabetes*. Obtenido de Universidad de Guadalajara: <https://udg.mx/es/noticia/en-mexico-la-mayoria-de-fallecidos-por-covid-19-tenian-diabetes>
- OMS. (2020). *Coronavirus*. Obtenido de https://www.who.int/es/health-topics/coronavirus#tab=tab_1
- R Core Team. (2020). R: A language and environment for statistical. *R Foundation for Statistical Computing*. Vienna, Austria. Obtenido de <https://www.R-project.org/>.

Autoevaluación

Propósito

Considero que el propósito principal de aprendizaje de este proyecto fue aplicar los conocimientos teóricos vistos en clase, en un problema real. Parte de esto incluye saber que herramientas estadísticas utilizar, como utilizarlas, y sobre todo elegir aquellas que son relevantes para una investigación (no solo usarlas por usar). Además, creo que era importante que lográramos llegar a conclusiones en el contexto del problema y siempre bien fundamentadas en el análisis

Estrategia

Durante el proyecto cambiamos algunas veces la estrategia a seguir como equipo. Al principio tuvimos que cambiar el enfoque del análisis porque no estaba dirigido a contestar la pregunta de investigación. En las actividades iniciales del Reto aprendimos de manera general algunas cosas que nos podían ser útiles, pero fue hasta las últimas dos actividades que me di cuenta del objetivo de aprendizaje, y busqué junto con mi equipo hacer un análisis relevante para resolver el problema. Cuando comenzamos a preparar la presentación también me di cuenta de algunas cosas que era necesario agregar para tener un trabajo completo.

Resultado

Considero que sí logré el propósito descrito anteriormente. Fue satisfactorio darme cuenta de que sí aprendí durante las clases y para el reto pude decidir la mejor manera de resolver un problema utilizando probabilidad y estadística. Específicamente algunas de las cosas más importantes que aprendí son las siguientes:

1. Aprovechar la tecnología para procesar e interpretar grandes cantidades de información. Al empezar esta materia, nunca creí que pudiera tener al alcance las

herramientas necesarias para realizar un análisis de una base de datos con 7 millones de entradas.

2. Analizar objetivamente información a través de la estadística y emitir juicios críticos que provienen de experimentación.
3. Para mí un aprendizaje indispensable fue el de argumentar correctamente y comunicar de forma coherente y estructurada los resultados finales de la actividad, tanto de manera escrita como oral.

Anexos

[Liga de drive a los archivos de análisis del reto](#)