

Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey



## Reporte Reto

¿Cómo ha sido el progreso de la pandemia?

Modelación estadística para la toma de decisiones

MA1001B

Grupo 203

Héctor David Bahena Garza | A01284661

Andrés Tarango Hernández | A01783336

Gonzalo Garza Moreno | A01284950

Alfonso Castro Camino | A01733148

2022-10-21

## Resumen

El proceso de vacunación en México empezó en diciembre del 2020, y hasta la fecha se ha reportado que el 68.01% de la población cuenta con la vacuna (Ritchie et. al, 2020). En este reporte buscamos medir la efectividad de estos esfuerzos comparando diferentes muestras: una al inicio de la pandemia, una durante el periodo de vacunación y otra más reciente. Se busca responder a la pregunta: ¿Hay una diferencia en el número de decesos por número de pacientes que ingresan respecto al tiempo? Los resultados nos dicen que hubo un decremento significativo en la tasa de muertes, tasa de hospitalizaciones y en la media de la edad de los casos totales, soportando la idea de que los esfuerzos de vacunación han sido efectivos.

## Introducción

La pandemia de COVID-19 es una pandemia actual declarada de tal manera por la Organización Mundial de la Salud el 11 de marzo de 2020. El virus SARS-CoV-2 ha afectado a más de 600 millones de personas alrededor del mundo en 260 países y territorios (Ritchie et. al, 2020). El virus llegó a México el 28 de febrero de 2020, y hasta la fecha se han confirmado más de 7 millones de casos, así como más de 300 mil defunciones a causa del virus dentro de la nación (Secretaría de Salud, 2022). Numerosos esfuerzos se han hecho a niveles municipales, estatales, nacionales y mundiales para prevenir el mayor contagio del virus, así como de reducir el número de fallecimientos en el país. Uno de esos esfuerzos ha sido el desarrollo de una vacuna. El proceso de vacunación empezó en diciembre del 2020, y hasta la fecha se ha reportado que el 68.01% de la población cuenta con la vacuna (Ritchie et. al, 2020). En este reporte buscamos medir la efectividad de estos esfuerzos, tomando en cuenta la tasa de muerte para diferentes muestras: una al inicio de la pandemia, y otra más reciente. De esta manera buscamos responder la pregunta: ¿Hay una diferencia en el número de decesos por número de pacientes que ingresan respecto al tiempo?

## Metodología

Hacer una muestra de las personas que ingresaron el segundo semestre de 2020, una del primer semestre de 2021 y otra del segundo semestre de 2021. Filtrar las muestras de acuerdo con la clasificación final para descartar los casos que no son SARS-CoV-2. Contabilizar el total de casos. Usar la fecha de defunción para saber si la persona falleció y contabilizar las defunciones. Contabilizar el número de hospitalizaciones. Crear proporciones con el total de casos, hospitalizaciones y defunciones para cada muestra. Finalmente se daría un intervalo de confianza para la diferencia de fallecimientos y de hospitalizaciones por casos de SARS-CoV-2 entre los periodos que estamos analizando con el fin de saber si la diferencia es significativa.

## SUBSETS DE PERIODOS

Primero se separó la muestra de 120000 observaciones en 3 muestras: de junio a diciembre 2020, de enero a junio 2021 y de junio a diciembre 2021. El muestreo fue aleatorio. Se obtuvo una muestra de 23807 observaciones para el primer periodo, 15872 para el segundo periodo y 27829 para el tercer periodo.

## UNA POBLACIÓN

Se exploraron las variables de cada muestra por separado.

### EDADES

#### INTERVALOS DE CONFIANZA

```
## Loading required package: lattice

##
## Attaching package: 'BSDA'

## The following object is masked from 'package:datasets':
##
##      Orange

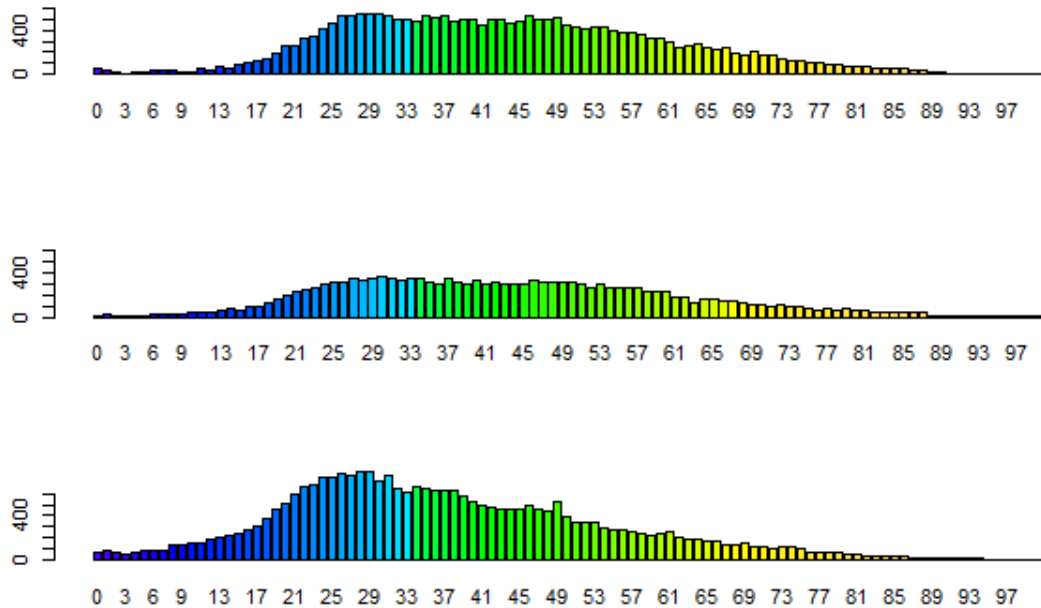
##
##  One-sample z-Test
##
## data:  e1
## z = 395.58, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  43.47646 43.90943
## sample estimates:
## mean of x
##  43.69295

##
##  One-sample z-Test
##
## data:  e2
## z = 313.86, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  42.93265 43.47222
## sample estimates:
## mean of x
##  43.20243

##
##  One-sample z-Test
##
## data:  e3
## z = 365.99, p-value < 2.2e-16
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  36.88889 37.28611
## sample estimates:
## mean of x
##  37.0875
```

Con esto se puede ver que las primeras 2 muestras tienen casi la misma media de edad. Hay un decremento aparente para la media de edad de la tercera muestra.

Gráfico de barras para las edades en primer, segundo y tercer periodo:



Estos gráficos nos permiten visualizar la reducción en la media de la edad

## TASAS DE MUERTES

### PRUEBA DE HIPÓTESIS

$H_0$ : La tasa de muertes en el primer periodo  $> 0.1$

$H_1$ : La tasa de muertes en el primer periodo  $< 0.1$

$$\alpha = 0.05$$

Regla de decisión: Se rechaza  $H_0$  si el valor-p  $< \alpha = 0.05$

## valor-p: 0.0008858784

valor-p = 1.284939e-05  $< \alpha$  0.05: Se rechaza  $H_0$ .

Se concluye que hay evidencia para demostrar que la tasa de muerte poblacional en el primer periodo es menor al 0.1.

## INTERVALOS DE CONFIANZA

Se obtuvieron intervalos de confianza para las proporciones de decesos por total de casos en las 3 muestras.

```
##
## 1-sample proportions test without continuity correction
##
## data:  s1_muertes out of s1_casos, null probability 0.5
## X-squared = 15703, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.09028156 0.09769338
## sample estimates:
##           p
## 0.09392196

##
## 1-sample proportions test without continuity correction
##
## data:  s2_muertes out of s2_casos, null probability 0.5
## X-squared = 10931, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.08081530 0.08949639
## sample estimates:
##           p
## 0.08505544

##
## 1-sample proportions test without continuity correction
##
## data:  s3_muertes out of s3_casos, null probability 0.5
## X-squared = 23360, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.03960728 0.04431665
## sample estimates:
##           p
## 0.04189874
```

## TASA DE HOSPITALIZACIÓN

### PRUEBA DE HIPÓTESIS

H<sub>0</sub>: La tasa de hospitalización en el primer periodo > 0.2

H<sub>1</sub>: La tasa de hospitalización en el primer periodo < 0.2

$$\alpha = 0.05$$

Regla de decisión: Se rechaza  $H_0$  si el valor-p <  $\alpha = 0.05$

```
## valor-p: 0.0001890796
```

valor-p =  $8.415387 \times 10^{-7}$  <  $\alpha = 0.05$ : Se rechaza  $H_0$ .

Se concluye que hay evidencia para demostrar que la tasa de hospitalización poblacional en el primer periodo es menor al 0.2.

### INTERVALOS DE CONFIANZA

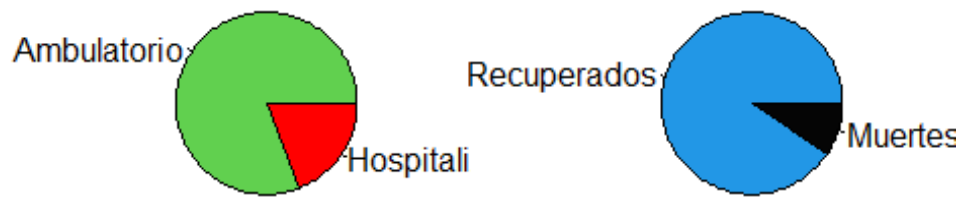
```
##
## 1-sample proportions test without continuity correction
##
## data:  s1_hosps out of s1_casos, null probability 0.5
## X-squared = 9105.2, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.1858431 0.1958251
## sample estimates:
##           p
## 0.1907842

##
## 1-sample proportions test without continuity correction
##
## data:  s2_hosps out of s2_casos, null probability 0.5
## X-squared = 7023.1, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.1616743 0.1732901
## sample estimates:
##           p
## 0.1674017

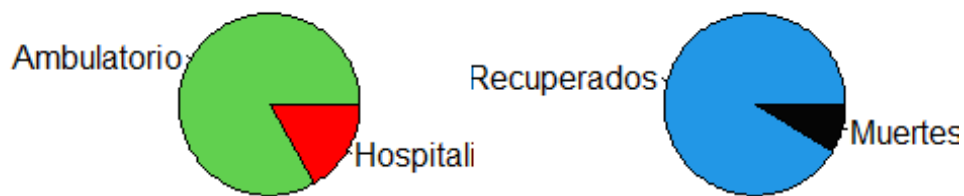
##
## 1-sample proportions test without continuity correction
##
## data:  s3_hosps out of s3_casos, null probability 0.5
## X-squared = 18038, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.09402324 0.10099248
## sample estimates:
##           p
## 0.0974523
```

### Pie Charts

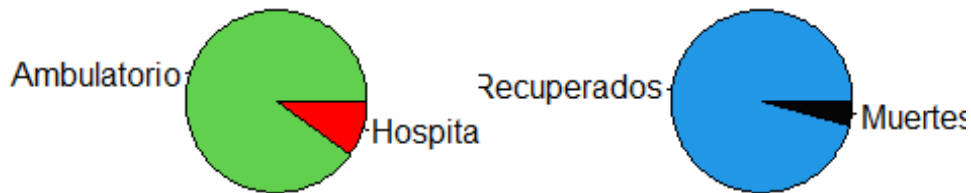
Periodo 1:



Periodo 2:



Periodo 3:



A partir de esta exploración, se aprecia un decremento en la cantidad relativa de hospitalizaciones y de muertes conforme avanza el tiempo. Esto se argumentará en las hipótesis de diferencia de proporciones para dos poblaciones a continuación.

## DOS POBLACIONES

### DIFERENCIA EDADES (PRIMER PERIODO - SEGUNDO PERIODO)

$H_0$ :  $e_1 = e_2$ : No existe una diferencia entre el promedio de edad en los diferentes periodos.

$H_1$ :  $e_1 \neq e_2$ : Sí existe una diferencia

$$\alpha = 0.05$$

Regla de decisión: Se rechaza  $H_0$  si el valor-p <  $\alpha = 0.05$

```
##  
## Two-sample z-Test  
##  
## data: e1 and e2  
## z = 2.7794, p-value = 0.005447  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## 0.1446107 0.8364203  
## sample estimates:  
## mean of x mean of y  
## 43.69295 43.20243
```

Como son muestras grandes, se usó la distribución normal Z (TCL).

El cero no está dentro del intervalo de confianza, por lo que se infiere que sí existe una diferencia.

valor-p = 0.00042 <  $\alpha$  0.05: Se rechaza  $H_0$ .

Se concluye que hay evidencia para demostrar que el promedio de edad a nivel nacional de aquellos reportados como casos positivos de COVID-19 disminuyó.

## DIFERENCIA TASAS DE MUERTES (PRIMER PERIODO - SEGUNDO PERIODO)

$H_0$ : tasa de muerte en periodo 1 = tasa de muerte en periodo 2: No existe una diferencia entre la tasa de muerte en los diferentes periodos.

$H_1$ : tasa de muerte en periodo 1  $\neq$  tasa de muerte en periodo 2: Sí existe una diferencia.

$$\alpha = 0.05$$

Regla de decisión: Se rechaza  $H_0$  si el valor-p <  $\alpha = 0.05$

```
##  
## 2-sample test for equality of proportions without continuity correcti  
on  
##  
## data: c(s1_muertes, s2_muertes) out of c(s1_casos, s2_casos)  
## X-squared = 9.1069, df = 1, p-value = 0.002547  
## alternative hypothesis: two.sided  
## 95 percent confidence interval:  
## 0.003159804 0.014573220  
## sample estimates:  
## prop 1 prop 2  
## 0.09392196 0.08505544
```



La diferencia de proporciones también se distribuye como una normal cuando la muestra es lo suficientemente grande, tal que la desviación estándar es la raíz cuadrada de la suma de los estimadores de proporción por su complemento entre n.

R hace este proceso automáticamente.

El cero no está dentro del intervalo de confianza, por lo que se infiere que sí existe una diferencia.

valor-p = 0.01576 < alfa = 0.05: Se rechaza H<sub>0</sub>.

Se concluye que hay evidencia para demostrar que la tasa de mortalidad por COVID-19 disminuyó.

### DIFERENCIA TASAS DE HOSPITALIZACIÓN (PRIMER PERIODO - SEGUNDO PERIODO)

H<sub>0</sub>: tasa de hospitalización en periodo 1 = tasa de hospitalización en periodo 2: No existe una diferencia entre la tasa de hospitalización en los diferentes periodos.

H<sub>1</sub>: tasa de hospitalización en periodo 1 ≠ tasa de hospitalización en periodo 2: Sí existe una diferencia.

$$\alpha = 0.05$$

Regla de decisión: Se rechaza H<sub>0</sub> si el valor-p < alpha = 0.05

```
##
## 2-sample test for equality of proportions without continuity correction
##
## data: c(s1_hosps, s2_hosps) out of c(s1_casos, s2_casos)
## X-squared = 35.058, df = 1, p-value = 3.2e-09
## alternative hypothesis: two.sided
## 95 percent confidence interval:
##  0.01572452 0.03104050
## sample estimates:
##      prop 1      prop 2
## 0.1907842 0.1674017
```

El mismo proceso se aplicó para la diferencia de proporciones, pero ahora con la variable de hospitalizaciones.

El cero no está dentro del intervalo de confianza, por lo que se infiere que sí existe una diferencia.

valor-p = 1.633e-08 < alfa = 0.05: Se rechaza H<sub>0</sub>.

Se concluye que hay evidencia para demostrar que la tasa de hospitalización por COVID-19 disminuyó.

## DIFERENCIA TASAS DE MUERTES EN DOS ESTADOS DE LA REPÚBLICA (NUEVO LEÓN - PUEBLA)

H\_0: tasa de muerte en Nuevo León = tasa de muerte en Puebla: No existe una diferencia entre la tasa de muerte de Nuevo León y Puebla.

H\_1: tasa de muerte en Nuevo León  $\neq$  tasa de muerte en Puebla: Sí existe una diferencia.

$$\alpha = 0.05$$

Regla de decisión: Se rechaza H\_0 si el valor-p < alpha = 0.05

```
##  
## 2-sample test for equality of proportions without continuity correcti  
on  
##  
## data: c(NL_muertes, PU_muertes) out of c(NL_casos, PU_casos)  
## X-squared = 164.58, df = 1, p-value < 2.2e-16  
## alternative hypothesis: two.sided  
## 95 percent confidence interval:  
## -0.07154513 -0.05115850  
## sample estimates:  
## prop 1 prop 2  
## 0.03109170 0.09244352
```

Por último, se obtuvo el intervalo de confianza para la diferencia de proporciones, regresando a la variable de muertes pero comparando dos entidades en lugar de los periodos de tiempo que se muestrearon originalmente.

El cero no está dentro del intervalo de confianza, por lo que se infiere que sí existe una diferencia.

valor-p = 2.2e-16 < alfa = 0.05: Se rechaza H\_0.

Se concluye que hay evidencia para demostrar que la tasa de muertes en Puebla es mayor que la tasa de muertes en Nuevo León.

## INDEPENDENCIA

Como un complemento al análisis de diferencia en las tasas de hospitalizaciones, deseamos probar si la obesidad y la hospitalización no eran independientes como lo indicaban expertos en salud.

H\_0: La obesidad y la hospitalización de un paciente son independientes.

H\_1: No son independientes.

$$\alpha = 0.05$$

```
## Obesidad

##
##      1      2      98
## 11802 107875    323

##
## Tipo Paciente

##
##      1      2
## 107764 12236
```

Nueva base de datos sin los 98

```
## Obesidad

##
##      1      2
## 11802 107875

##
## Tipo Paciente

##
##      1      2
## 107483 12194
```

### Tabla bivariada

```
##
##      SÍ      NO
## SÍ  9453  2349
## NO  98030 9845

##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  Ta
## X-squared = 1349, df = 1, p-value < 2.2e-16
```

Como el valor  $p = 0.0000 < 0.05$  se rechaza  $H_0$ . Por tanto, hay evidencia para decir que no son independientes.

### ANOVA

```
##
##              Df    Sum Sq Mean Sq F value Pr(>F)
as.Date(M$FECHA_INGRESO)    1    604165    604165    2144 <2e-16 ***
Residuals              119998  33811163      282
```

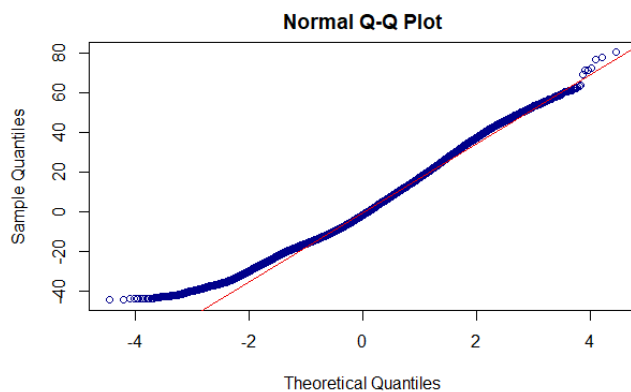
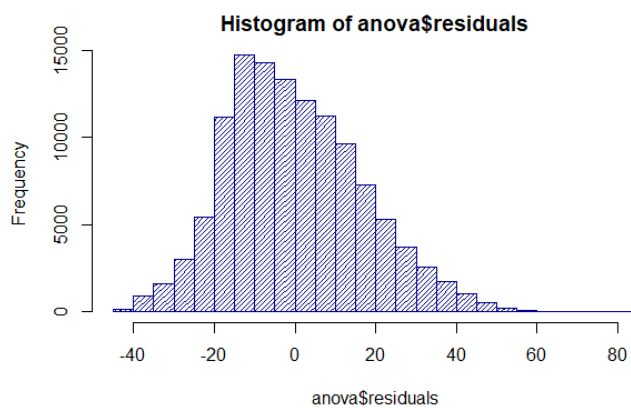
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Se realizó un ANOVA para la edad respecto a la fecha de ingreso. El valor-p que nos arrojó es mucho menor a alfa, por lo que se puede decir que la fecha de ingreso tiene una fuerte influencia en la edad de los pacientes. Esto soporta la diferencia de medias que se observó anteriormente entre los distintos periodos.

También se verificaron los supuestos de ANOVA para obtener el siguiente resultado

Normalidad



```
##
```

```
## Anderson-Darling normality test
```

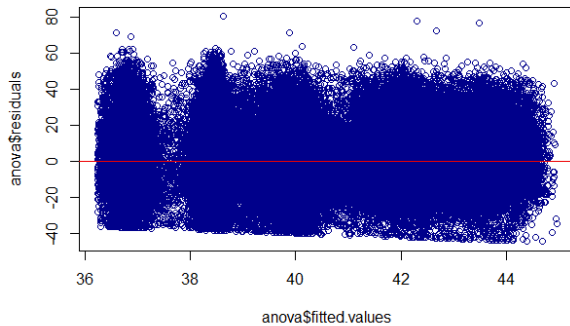
```
##
```

```
## data: anova$residuals
```

```
## A = 451.41, p-value < 2.2e-16
```

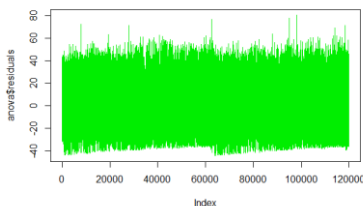
La gráfica Q-Q y el test de Anderson-Darling nos confirman que hay normalidad en los residuos.

## Homocedasticidad



Parece haber varianza constante en los residuos. Se observa como encajan dentro de 2 rectas horizontales alrededor de la media, aunque hay un ligero sesgo en la parte izquierda.

## Independencia de los errores



No se observa una tendencia en los residuos.

Se cumplen los supuestos del análisis de varianza, por lo que el resultado es confiable

## Discusión y Conclusiones

A través de la metodología seguida, se observó que hubo una reducción en la media de las edades. Esto se comprobó usando pruebas de hipótesis, y posteriormente fue confirmado por el análisis de varianza. De acuerdo con el análisis, la fecha de ingreso tiene una fuerte influencia en la edad del paciente.

Con base en los resultados obtenidos, se puede llegar a la conclusión que es probable que sí haya habido una mejoría en las tasas de hospitalización y de muertes, con una certeza del 95%. Ambas tasas se redujeron dramáticamente para el tercer periodo, que fue ocurrió cuando estaba por completarse la vacunación en México. Esto soporta la idea de que los esfuerzos de vacunación fueron efectivos para reducir la gravedad de la pandemia.

Otros hechos de los que nos percatamos, aunque no estaban estrechamente relacionados a la pregunta de investigación, fue que algunas entidades presentaban una menor disminución en las tasas de hospitalización y fallecimientos. Puede ser debido a campañas de vacunación menos exitosas o algunos factores desconocidos. También corroboramos lo que afirmaban las autoridades en salud, que la obesidad es un factor de riesgo para la hospitalización.

La mejora observada en las tasas de hospitalización y fallecimientos se puede asociar a los esfuerzos mayores realizados por prevención y cuidados de los pacientes, notablemente la aplicación de la vacuna, ya que las fechas en las que se observó un gran cambio fueron al inicio de la campaña de vacunación y cuando se completó la vacunación por edades.

Sin embargo, parte de los efectos observados se pueden deber a cambios que no tomamos en consideración, como fenómenos sociales. Por ejemplo, las olas que hubo debido a las vacaciones y el hecho de que los más jóvenes eran los más probables en salir y viajar durante esa época. Otra limitación es que se asumió normalidad para las medias y las proporciones por tratarse de muestras grandes. Los resultados podrían ser distintos si se usara un nivel de significancia más estricto, como  $\alpha=0.01$ , para reducir la probabilidad de cometer error tipo 1.

Algunas preguntas que se suscitan, pero han quedado fuera del alcance de este análisis, son: ¿Qué ocurriría si se toman muestras para periodos de tiempo más cortos? ¿Se observará la misma tendencia si se extiende el análisis a fechas más recientes, o la reducción en las tasas se detendrá debido a que la mayor parte de la población ya está vacunada? ¿Las diferencias que se observaron entre entidades corresponden a diferencias en la cantidad de vacunas aplicadas? Respuestas a estas preguntas nos proporcionarían todavía más información sobre la efectividad de la vacuna y sobre el comportamiento que se esperaría para el futuro desarrollo de la pandemia.

## Referencias

Hannah Ritchie, Edouard Mathieu, Lucas Rodés-Guirao, Cameron Appel, Charlie Giattino, Esteban Ortiz-Ospina, Joe Hasell, Bobbie Macdonald, Diana Beltekian, & Max Roser (2020). Coronavirus Pandemic (COVID-19). Our World in Data.

Secretaría de Salud, (2022). Comunicado Técnico Semanal (COVID-19). Recuperado de <https://www.gob.mx/salud/prensa/comunicado-tecnico-semanal-covid-19-316224?idiom=es>

## Autoevaluación

### Propósito:

Me parece que el propósito de la actividad era no quedarnos solamente con el conocimiento teórico sobre lo aprendido acerca de estadística, sino saber cómo aplicarlo y resolver los problemas que se presentan en esta aplicación. Creo que el profesor quería que aprendiéramos a utilizar estas herramientas para analizar información y poder interpretar a qué conclusiones nos lleva la evidencia.

### Estrategia:

El plan fue proponer una pregunta que detonara la investigación y trabajar colaborativamente para procesar los datos y llegar a una conclusión. En nuestro equipo, el plan cambió ligeramente. En un inicio solo nos preguntamos sobre la diferencia en los decesos entre el semestre que todavía no se aplicaban las vacunas y el semestre en el que se comenzaron a aplicar, pero después extendimos nuestro análisis para otras variables como hospitalización y edad, además de agregar una muestra para el tercer periodo. A pesar de este cambio, siempre tuve claro lo que se tenía que hacer y la clase nos proporcionó con lo necesario para llevar a cabo el proyecto. Una vez que lo terminamos fue cuando me di cuenta de los principales objetivos de aprendizaje. Solucionar los problemas que se nos presentaron y aplicar los métodos que aprendí en las últimas 5 semanas me ayudaron para fortalecer este conocimiento y desarrollar nuevas habilidades.

### Resultado:

Siento que logré el propósito de aprendizaje, además me ha surgido una mayor curiosidad por el campo de la estadística a partir de este proyecto. Lo más importante que aprendí fue la resolución de problemas en programación, la aplicación de métodos estadísticos para extraer información sobre un conjunto de datos y el pensamiento inductivo que nos conduce a la creación de conocimiento.

## Anexo

<https://drive.google.com/drive/folders/1BcrgY80gMkGAfXnMEROYV1ifnPApwhWI?usp=sharing>