

10/21/2022

# El COVID-19 en México

Modelación Estadística para la  
Toma de Decisiones

Annette Pamela Ruiz Abreu - A01423595  
José Abraham Martínez Licona - A01368551  
Juan Pablo Solís Ruiz - A01067387  
Mayra Stefany Gómez Triana- A01625609

---

## Índice

Resumen	2
Introducción	3
Objetivo	4
Pregunta detonadora	5
Base de datos	6
Análisis	7
Exploración inicial de variables de interés	7
Estudio de proporciones y frecuencias	9
Estudio por estados	10
Tablas bivariadas	17
Análisis ANOVA	20
Conclusión	24
Catálogo de Entidades	25
Referencias	26

## Resumen

Este reporte analiza los casos reportados de COVID-19 en México para intentar encontrar *insights* que permitan un mejor entendimiento del virus y de su propagación en México. Para hacerlo, se utilizó el lenguaje de programación R y varias herramientas estadísticas como las medidas de tendencia central, gráficas, intervalos de confianza, pruebas de hipótesis y análisis ANOVA. Usando estas técnicas se encontraron factores que afectan la mortalidad, el estado con mayor proporción de muertes y el sexo que más muere.

## Introducción

El 31 de diciembre de 2019 en Wuhan, China se reportó el primer caso de un virus que cambiaría el mundo para siempre: COVID-19. El coronavirus es una gran familia de virus conocidos por causar enfermedades que van desde un resfriado común hasta manifestaciones clínicas más severas como las observadas en el Síndrome respiratorio por el coronavirus de Oriente Medio (MERS) y el Síndrome respiratorio agudo grave (SARS). Expertos mundiales, gobiernos y organizaciones han trabajado arduamente para ampliar rápidamente los conocimientos científicos sobre este nuevo virus, rastrear su propagación y virulencia y asesorar a los países y las personas sobre las medidas para proteger la salud y prevenir la propagación del brote. En México ha habido un rápido aumento en la investigación, en respuesta al brote de COVID-19. Durante este período inicial, la investigación publicada exploró principalmente la epidemiología, las causas, la manifestación clínica y el diagnóstico, así como la prevención y el control del nuevo coronavirus. Inicialmente, hubo más artículos de investigación centrados en las causas del brote, pero los estudios sobre prevención y control aumentaron gradualmente con el tiempo.

## Objetivo

El análisis objetivo de información confiable permite entender críticamente el alcance, a veces insospechado, de las decisiones tomadas por los líderes responsables, pero también orienta para tomar las mejores decisiones. El reto consiste en procesar, entender críticamente y orientar hacia las mejores decisiones sobre uno de los grandes problemas de nuestro tiempo: la pandemia del Covid-19 a partir de la información abierta y continuamente actualizada de la Secretaría de Salud en México. El objetivo del reporte es contestar la pregunta detonadora que se presenta a continuación y encontrar información relevante sobre el problema mediante técnicas estadísticas como las pruebas de hipótesis y las tablas bivariadas.

### **Pregunta Detonadora**

De los pacientes registrados, ¿quiénes son los que más fallecen, los hombres o las mujeres?  
¿En qué estado hubo más muertes por COVID-19 en México? ¿La edad, la entubación, la obesidad y la diabetes son factores para la defunción de los pacientes?

## Base de Datos

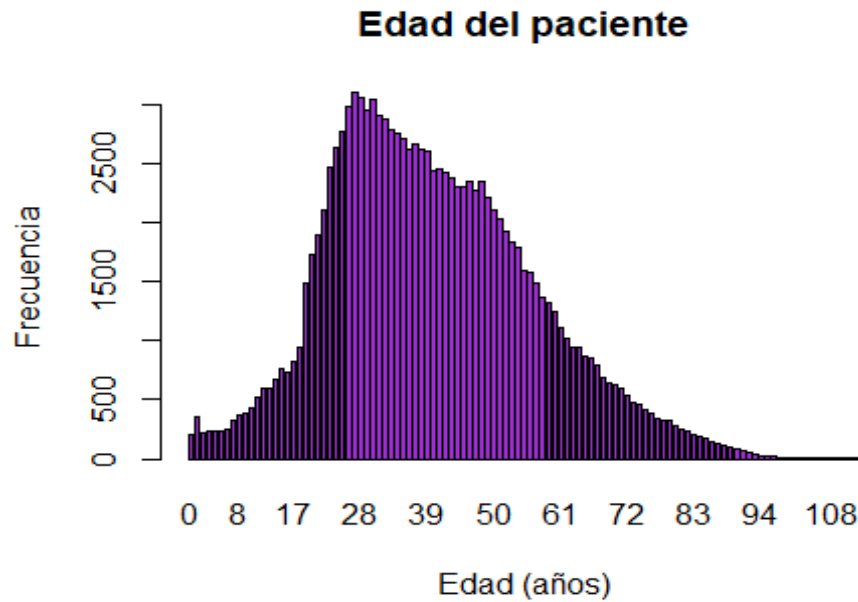
Se utilizó una base de datos de la Secretaría de Salud mexicana que contenía los registros de todas las personas con COVID-19 confirmado desde el 14 de septiembre de 2022. Se tomó una muestra de 120 000 datos, respetando la proporción de mujeres y hombres de la población. La base de datos tenía las siguientes columnas: fecha\_actualización, id\_registro, origen, sector, entidad\_um, entidad\_nac, entidad\_res, municipio\_res, tipo\_paciente, fecha\_ingreso, fecha\_sintomas, fecha\_def, intubado, neumonía, edad, nacionalidad, embarazo, habla\_lengua\_indig, indígena, diabetes, epoc, asma, inmusupr, hipertensión, otra\_com, cardiovascular, obesidad, renal\_crónica, tabaquismo, otro\_caso, toma\_muestra\_lab, toma\_muestra\_antígeno, resultado\_antígeno, clasificación\_final, migrante, país\_nacionalidad, país\_origen, uci, sexo.

## Variables de Interés

- Sexo
- Edad
- Fecha de defunción
- Entidad Medica (Entidad UM)
- Días hospitalizado hasta defunción
- Intubado

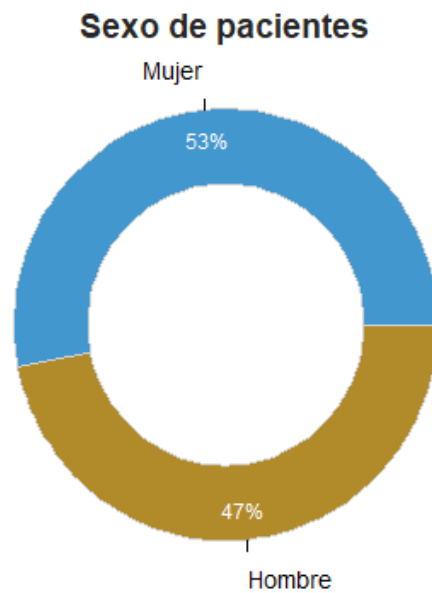
## Análisis

### Exploración inicial de variables de interés



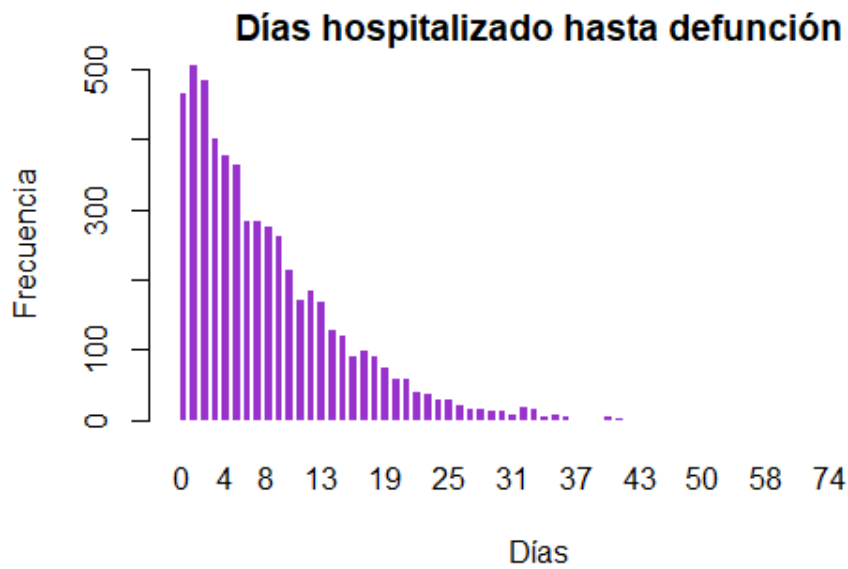
#### Medidas de Tendencia Central

- Media: 39.88086
- Mediana: 38
- Moda: 27
- Rango: 0-120
- Varianza: 285.8338
- Desviación: 16.9066



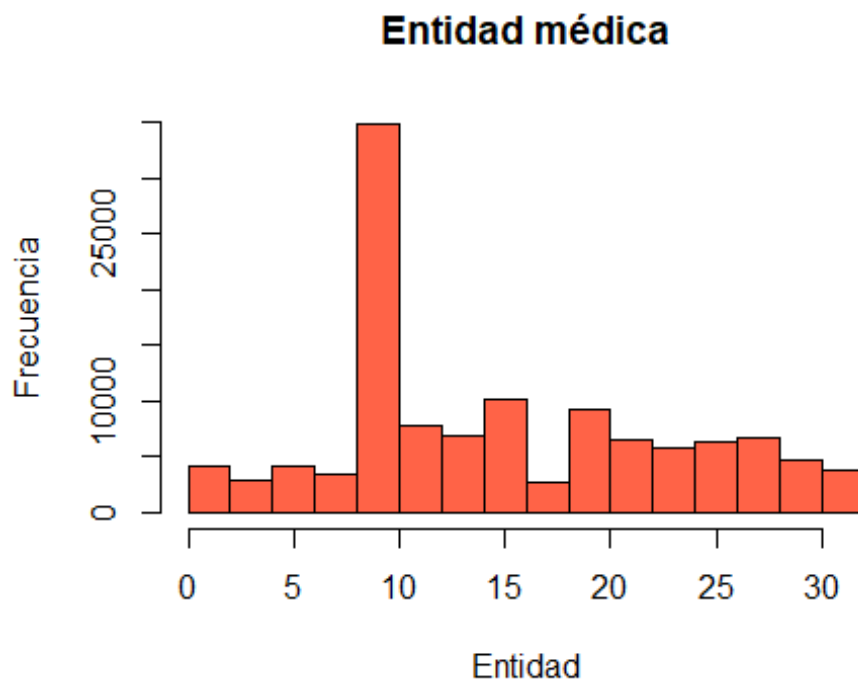
	Mujer	Hombre	Total
Frecuencias	63780	56220	120000
Proporciones	0.531	0.469	1





#### Medidas de Tendencia Central

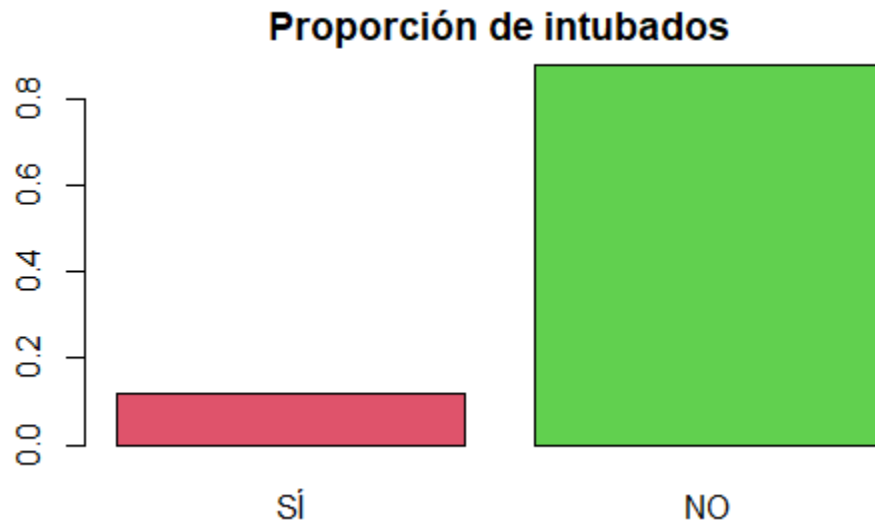
- Media: 8.499373
- Mediana: 6
- Moda: 1
- Rango: 0-463
- Varianza: 124.0555
- Desviación: 11.138



- Moda: 9 (C)

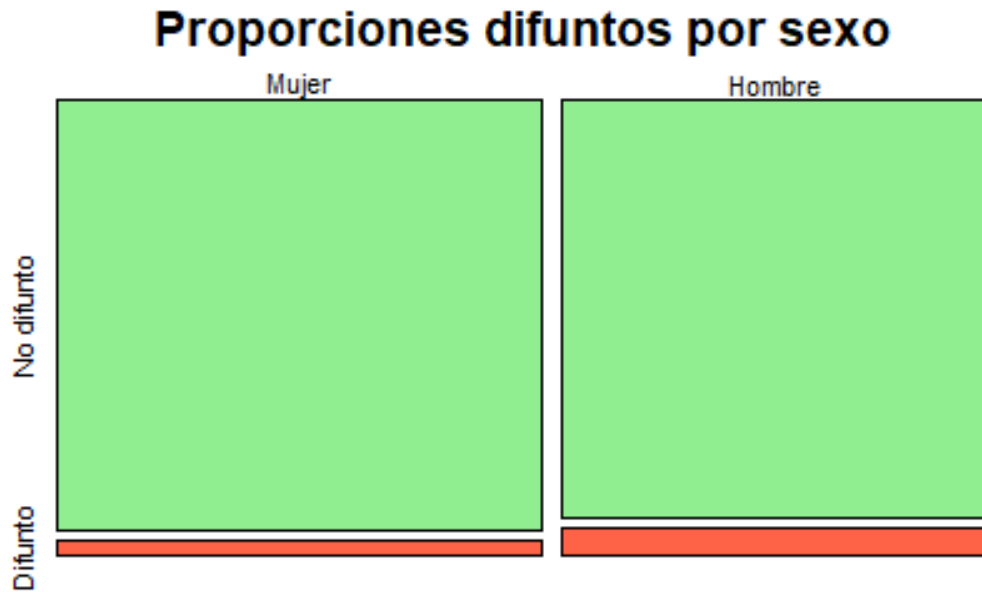
Revisar catálogo de entidades (página 10)

## Estudio de proporciones y frecuencias



La base de datos original no tenía una columna que especificara si el paciente murió o no; por ende, para facilitar el análisis, se creó una columna nueva en donde se supuso que los pacientes que no tenían fecha de defunción no murieron, y los demás sí. En esta columna nueva 0 significa que el paciente no murió y 1 significa que sí murió.

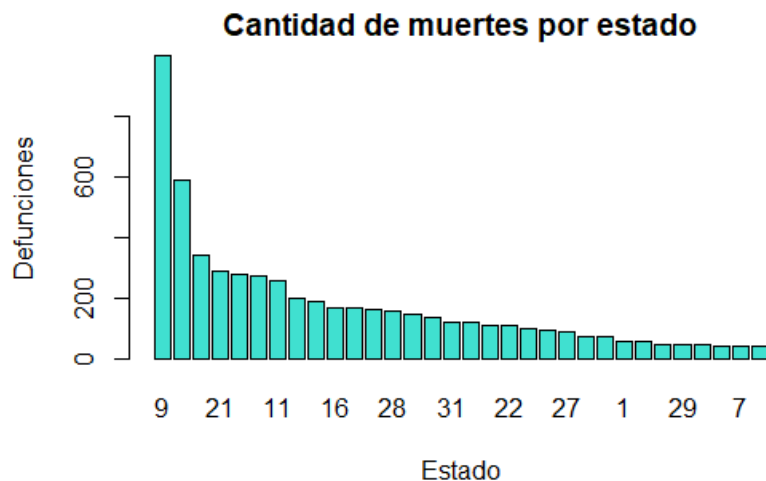




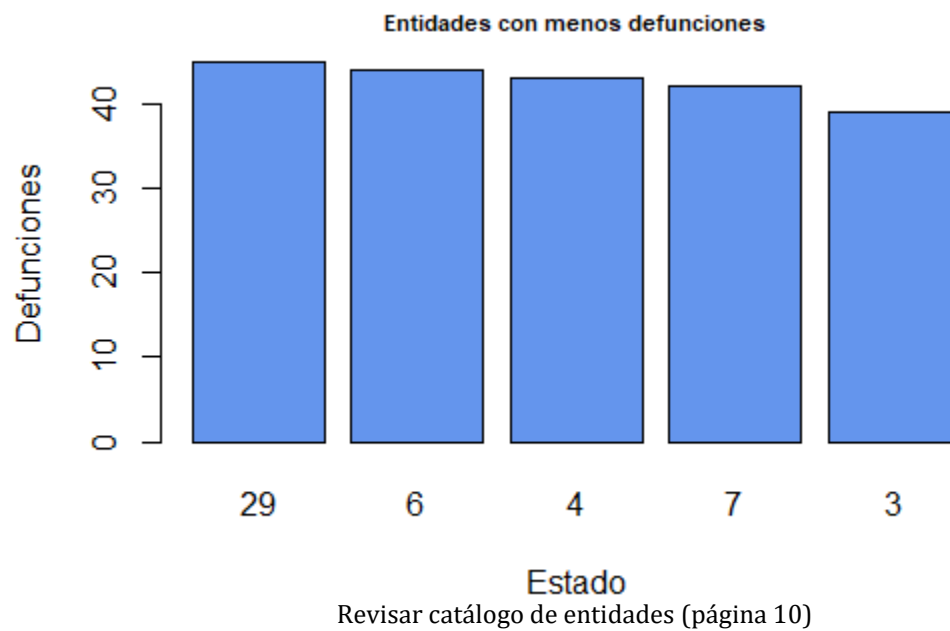
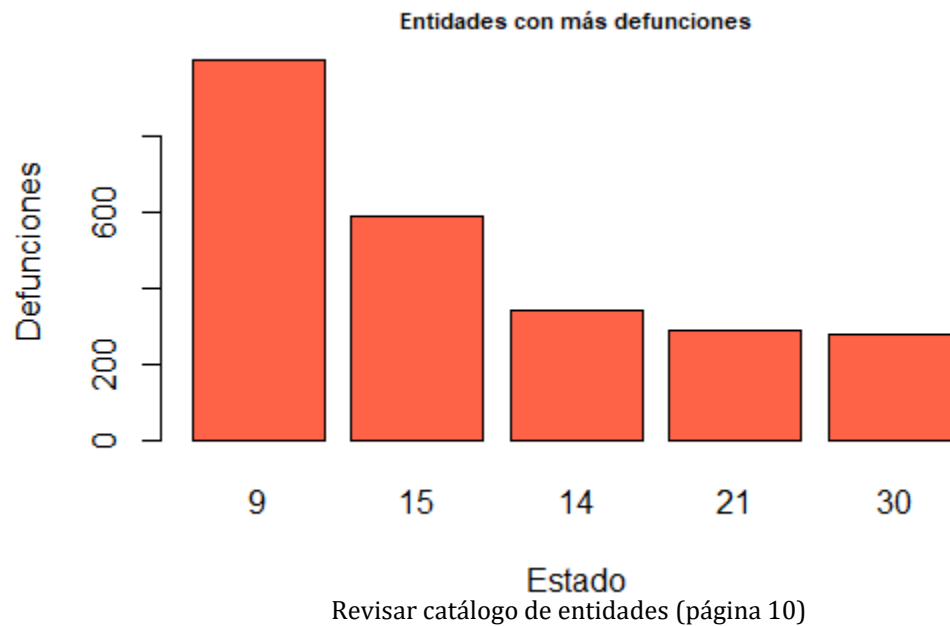
En esta gráfica podemos observar que aunque en la muestra hay más mujeres que hombres, la proporción de hombres difuntos es mayor a la de mujeres. Por ende, podemos concluir que los hombres tienen una mayor tasa de mortalidad por COVID-19 que las mujeres.

### Estudio por estados

Inicialmente, se analizó la cantidad de muertes (la frecuencia) en cada estado. Esto no arrojó como resultado que los estados con mayor cantidad de muertes eran: la ciudad de México, el estado de México y Jalisco.



### Top 5 entidades médicas con más y menos defunciones

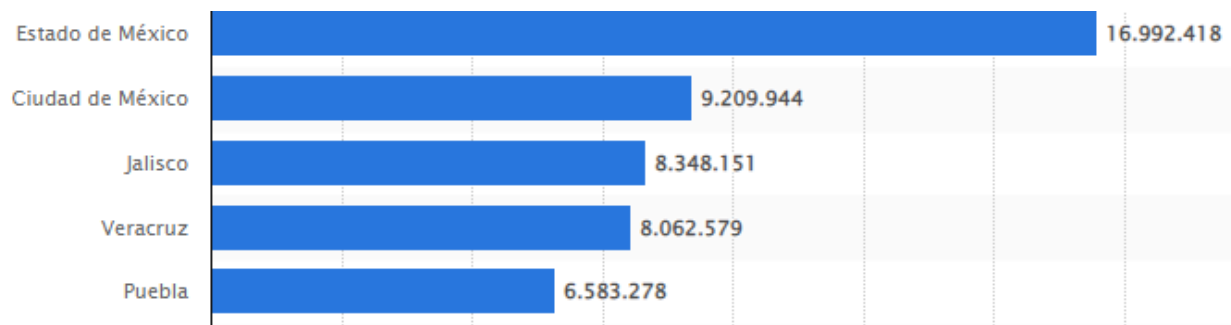


Con base en estas gráficas, concluiríamos que los estados con más muertes son la ciudad de México, el estado de México y Jalisco, y haríamos un análisis exhaustivo para encontrar la razón de esto. Sin embargo, estos primeros resultados son engañosos porque al revisar los

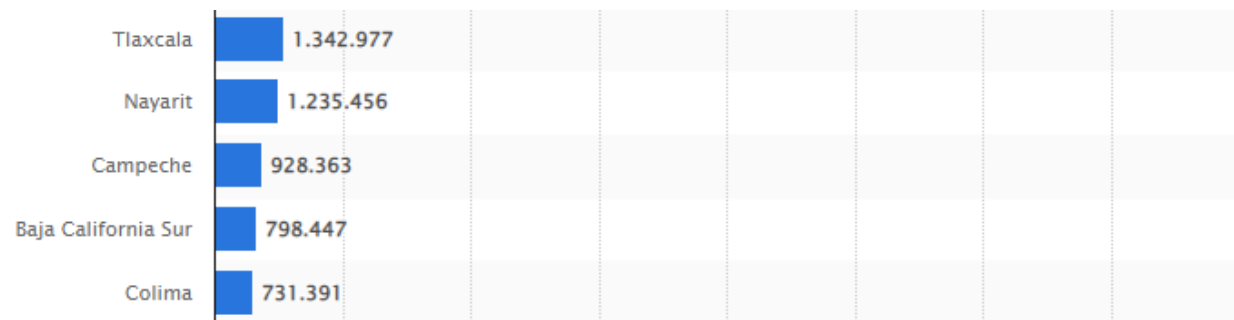
datos del Instituto Nacional de Estadística y Geografía (INEGI) encontramos que los estados con más defunciones coinciden con los estados con mayor población.

### Número de habitantes en México en 2020 por estado

#### 5 estados con más habitantes



#### 5 estados con menos habitantes



Para comprobar si el estado con más cantidad de muertos es el estado con mayor proporción de muertos, se realizó un intervalo de confianza del 99 % de la diferencia de proporciones entre los dos estados con mayor cantidad de muertos y una prueba de hipótesis.

#### *Intervalo de confianza de defunción en los estados con más defunciones*

```
prop.test(c(f1, f2), c(n1, n2), conf.level=NC, correct=FALSE)
```

2-sample test for equality of proportions without continuity correction

data: p out of n

X-squared = 307.05, df = 1, p-value < 0.000000000000000022

alternative hypothesis: two.sided

99 percent confidence interval:

-0.04852113 -0.03330539

sample estimates:

prop 1 prop 2

0.02974542 0.07065868

[1] -0.04852113 -0.03330539

Se estima, con un nivel de confianza del 99 %, que la diferencia entre la proporción 1 (proporción de muertos en la ciudad de México) y la proporción 2 (proporción de muertos en el estado de México) está en el siguiente intervalo [-0.0485 , -0.0333].

### *Prueba de hipótesis de las proporciones de los estados con más defunciones*

$H_0$ : La proporción de muertos en el Distrito Federal es igual a la proporción de muertos en el Estado de México.

$H_0$ : pDF = pMC

$H_1$ : La proporción de muertos en el Distrito Federal es mayor que la proporción de muertos en el Estado de México.

$H_1$ : pDF != pMC

alfa = 0.01

p1 = f1/n1

p2 = f2/n2

pp = (n1\*p1+n2\*p2)/(n1+n2)

Zp = (p1-p2)/sqrt(pp\*(1-pp)\*(1/n1+1/n2))

Z = qnorm(alfa/2)

Como el estadístico  $Z^*$  es menor que el valor frontera Z, se rechaza la hipótesis nula.

Como el valor-p es menor o igual que alfa, se rechaza la hipótesis nula.

## Conclusión

Con base en los resultados del intervalo de confianza y las pruebas de hipótesis, podemos concluir que las proporciones de muertes en la ciudad de México y el estado de México son diferentes, gracias a que rechazamos la hipótesis nula y porque el cero no está dentro del intervalo de confianza de la diferencia de proporciones entre DF y MC. Además, como el intervalo es negativo, se estima que la proporción de muertos en la ciudad de México en realidad es menor que la proporción de muertos en el estado de México. Es decir, aunque la ciudad de México tiene más cantidad de muertes, no tiene la mayor proporción.

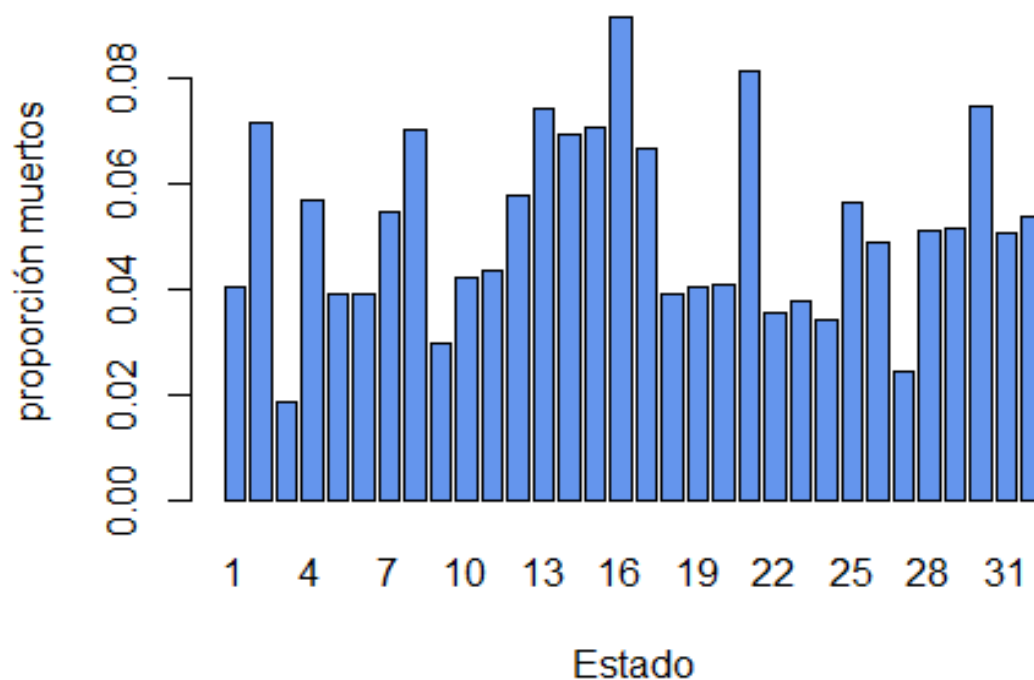
Como hemos observado, el hecho de que un estado tenga mayor frecuencia de defunciones que otro no significa que tenga una mayor proporción de muertes. Para encontrar el estado con mayor proporción de muertes, se ha hecho la siguiente tabla y gráfica:

*Tabla de proporciones*

ESTADO	VIVOS	MUERTOS	OBESOS	NO.OBESOS	DIABETICOS	NO.DIABETICOS	INTUBADOS	NO.INTUBADOS
1	0.959862	0.040138	0.085121	0.914879	0.078201	0.921799	0.014533	0.086505
2	0.928701	0.071299	0.128122	0.871155	0.100977	0.898299	0.027868	0.101701
3	0.981429	0.018571	0.114762	0.884286	0.080476	0.918095	0.006667	0.039048
4	0.943197	0.056803	0.143989	0.856011	0.095112	0.904888	0.009247	0.100396
5	0.961133	0.038867	0.066535	0.923254	0.082016	0.908103	0.004941	0.085639
6	0.960993	0.039007	0.132092	0.867021	0.070922	0.928191	0.02305	0.054078
7	0.945596	0.054404	0.086788	0.913212	0.084197	0.915803	0.031088	0.068653
8	0.929786	0.070214	0.095713	0.903178	0.107908	0.890983	0.020325	0.109387
9	0.970255	0.029745	0.083698	0.911419	0.075689	0.91916	0.010124	0.05696
10	0.95814	0.04186	0.089147	0.9	0.068992	0.91938	0.010853	0.072093
11	0.956581	0.043419	0.109995	0.889835	0.089562	0.910267	0.005449	0.089222
12	0.942349	0.057651	0.099138	0.898707	0.116918	0.880927	0.025323	0.088901
13	0.925852	0.074148	0.119238	0.880261	0.099198	0.900301	0.018537	0.147295
14	0.930922	0.069078	0.124259	0.87431	0.10934	0.888412	0.012262	0.132638
15	0.929341	0.070659	0.094611	0.904671	0.08	0.918802	0.013293	0.144431
16	0.908739	0.091261	0.113938	0.884403	0.117257	0.880531	0.019358	0.139381
17	0.933706	0.066294	0.106071	0.893929	0.114445	0.885555	0.006281	0.122121
18	0.960996	0.039004	0.099585	0.899585	0.077178	0.921992	0.010788	0.079668
19	0.959657	0.040343	0.081572	0.917984	0.08438	0.915177	0.007832	0.086744
20	0.95925	0.04075	0.08965	0.91035	0.075387	0.924205	0.016707	0.074165
21	0.918697	0.081303	0.104533	0.894901	0.112181	0.887535	0.01728	0.154958
22	0.964675	0.035325	0.102674	0.896666	0.084846	0.914493	0.013206	0.075933
23	0.962366	0.037634	0.08172	0.91828	0.058065	0.941935	0.021505	0.060215
24	0.965983	0.034017	0.095798	0.8992	0.086543	0.905953	0.007004	0.054277
25	0.943954	0.056046	0.095109	0.903872	0.08322	0.915761	0.013587	0.094769
26	0.951497	0.048503	0.141617	0.858084	0.098503	0.900898	0.007186	0.090719
27	0.975933	0.024067	0.101406	0.898053	0.099513	0.899946	0.008112	0.042726
28	0.94908	0.05092	0.112024	0.887648	0.107096	0.89159	0.009198	0.094612
29	0.94863	0.05137	0.099315	0.900685	0.082192	0.917808	0.009132	0.105023
30	0.925421	0.074579	0.100508	0.896017	0.10425	0.893344	0.019781	0.141406
31	0.949704	0.050296	0.080727	0.918005	0.060017	0.938715	0.011412	0.084954
32	0.946205	0.053795	0.079587	0.920413	0.084009	0.914517	0.008843	0.086957



## Proporción de defunciones por estado



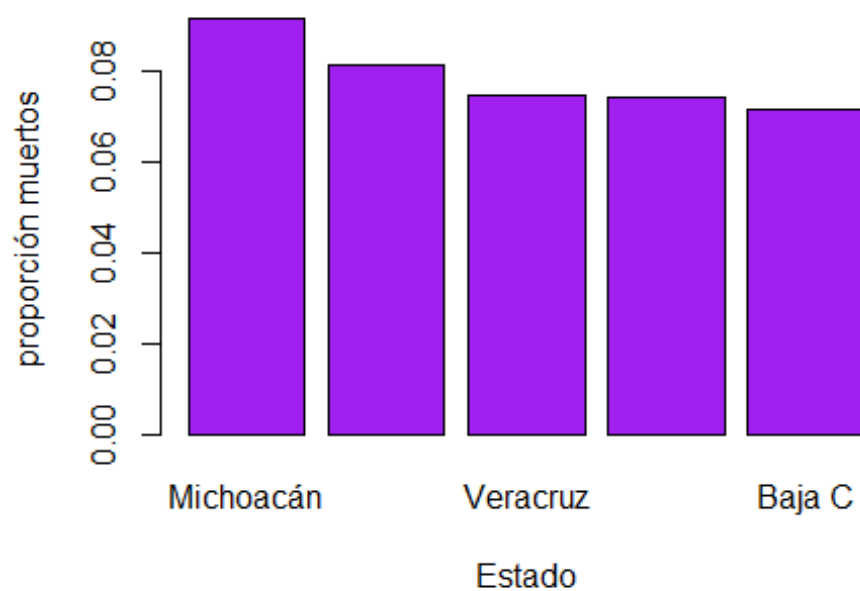
### Top 5 estados con mayor proporción de defunciones

ESTADO	NOMBRE	VIVOS	MUERTOS
16	Michoacán	0.908739	0.091261
21	Puebla	0.918697	0.081303
30	Veracruz	0.925421	0.074579
13	Hidalgo	0.925852	0.074148
2	Baja California	0.928701	0.071299

### Top 5 estados con menor proporción de defunciones

ESTADO	NOMBRE	VIVOS	MUERTOS
3	Baja California Sur	0.981429	0.018571
27	Tabasco	0.975933	0.024067
9	Ciudad de México	0.970255	0.029745
24	San Luis Potosí	0.965983	0.034017
22	Querétaro	0.964675	0.035325

## Proporción de defunciones por estado



## Tablas bivariadas

	Obeso	No obeso
Difunto	10448	1144
No difunto	103694	4424

	Intubado	No intubado
Difunto	209	1234
No difunto	6388	4056

	Diabético	No diabético
Difunto	8409	2003
No difunto	105701	3566

Usando la prueba de chi cuadrada, se realizaron 3 pruebas de hipótesis con un nivel de significancia del 0.05 para determinar si las variables de obesidad, intubación y diabetes afectan la mortalidad. Es decir, las pruebas de hipótesis se hicieron para observar si hay correlación entre las variables.

$\alpha = 0.05$

$p = \text{chisq.test(tabla)}$

$\text{decision} = (p\text{-value} < \alpha)$

### *Prueba obesidad*

$H_0$ : La obesidad y la muerte son independientes. Tener obesidad no afecta la probabilidad de morir.

$H_1$ : La obesidad y la muerte no son independientes. Tener obesidad sí afecta la probabilidad de morir.

### *Prueba intubación*

$H_0$ : La intubación y la muerte son independientes. Ser intubado no afecta la probabilidad de morir.

$H_1$ : La intubación y la muerte no son independientes. Ser intubado sí afecta la probabilidad de morir.

### *Prueba diabetes*

$H_0$ : La diabetes y la muerte son independientes. Tener diabetes no afecta la probabilidad de morir.

$H_1$ : La obesidad y la muerte no son independientes. Tener diabetes sí afecta la probabilidad de morir.

Los valores p de las 3 pruebas dieron menores que alfa (0.05); por ende, concluimos que la obesidad, diabetes e intubación afectan/aumentan la probabilidad de morir.

Con base en esto, se estudió la obesidad, diabetes e intubación para ver si estas variables explicaban por qué Michoacán es el estado con mayor proporción de muertes.

#### *Top 10 estados con mayor proporción de obesidad*

ESTADO	VIVOS	MUERTOS	OBESOS	NO.OBESC
4	0.943197	0.056803	0.143989	0.856011
26	0.951497	0.048503	0.141617	0.858084
6	0.960993	0.039007	0.132092	0.867021
2	0.928701	0.071299	0.128122	0.871155
14	0.930922	0.069078	0.124259	0.87431
13	0.925852	0.074148	0.119238	0.880261
3	0.981429	0.018571	0.114762	0.884286
16	0.908739	0.091261	0.113938	0.884403
28	0.94908	0.05092	0.112024	0.887648
11	0.956581	0.043419	0.109995	0.889835

#### *Top 10 estados con mayor proporción de diabéticos*

ESTADO	VIVOS	MUERTOS	DIABETICOS	NO.DIABETICOS
16	0.908739	0.091261	0.11725664	0.880530973
12	0.942349	0.057651	0.1169181	0.880926724
17	0.933706	0.066294	0.11444522	0.88555478
21	0.918697	0.081303	0.1121813	0.887535411
14	0.930922	0.069078	0.10933987	0.888412017
8	0.929786	0.070214	0.10790835	0.890983001
28	0.94908	0.05092	0.10709593	0.891590013
30	0.925421	0.074579	0.1042502	0.893344026
2	0.928701	0.071299	0.1009772	0.89829895
27	0.975933	0.024067	0.09951325	0.899945917

### Top 10 estados con mayor proporción de intubados

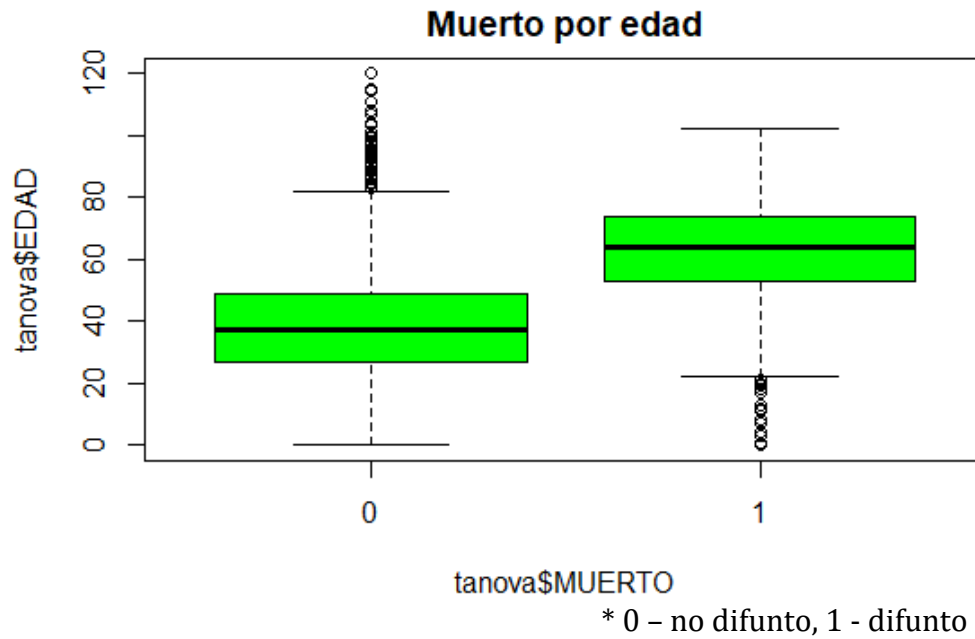
ESTADO	VIVOS	MUERTOS	INTUBADOS	NO.INTUBADOS
7	0.945596	0.054404	0.03108808	0.06865285
2	0.928701	0.071299	0.02786826	0.10170105
12	0.942349	0.057651	0.02532328	0.088900862
6	0.960993	0.039007	0.02304965	0.054078014
23	0.962366	0.037634	0.02150538	0.060215054
8	0.929786	0.070214	0.0203252	0.109386548
30	0.925421	0.074579	0.01978081	0.141406041
16	0.908739	0.091261	0.01935841	0.139380531
13	0.925852	0.074148	0.01853707	0.147294589
21	0.918697	0.081303	0.01728045	0.154957507

Como podemos observar, Michoacán es el estado con mayor proporción de personas diabéticos y es el octavo estado con mayor proporción de obesos e intubados. Aunque hay varios factores que contribuyen a la alta proporción de difuntos, podemos concluir gracias a las pruebas de hipótesis de las tablas bivariadas y a este análisis que Michoacán es el estado con más muertes gracias a la alta proporción de diabéticos, intubados y obesos.

### Análisis ANOVA

Para determinar si la edad es un factor que afecta en la mortalidad de los pacientes, se hizo un análisis ANOVA.

	No difunto	Difunto
Media	38.74891	63.05260
Mediana	27	67
Moda	37	64



$H_0$ : La edad y la muerte son independientes. La edad no afecta la probabilidad de morir.

$H_1$ : La edad y la muerte no son independientes. La edad sí afecta la probabilidad de morir.

### Modelo ANOVA

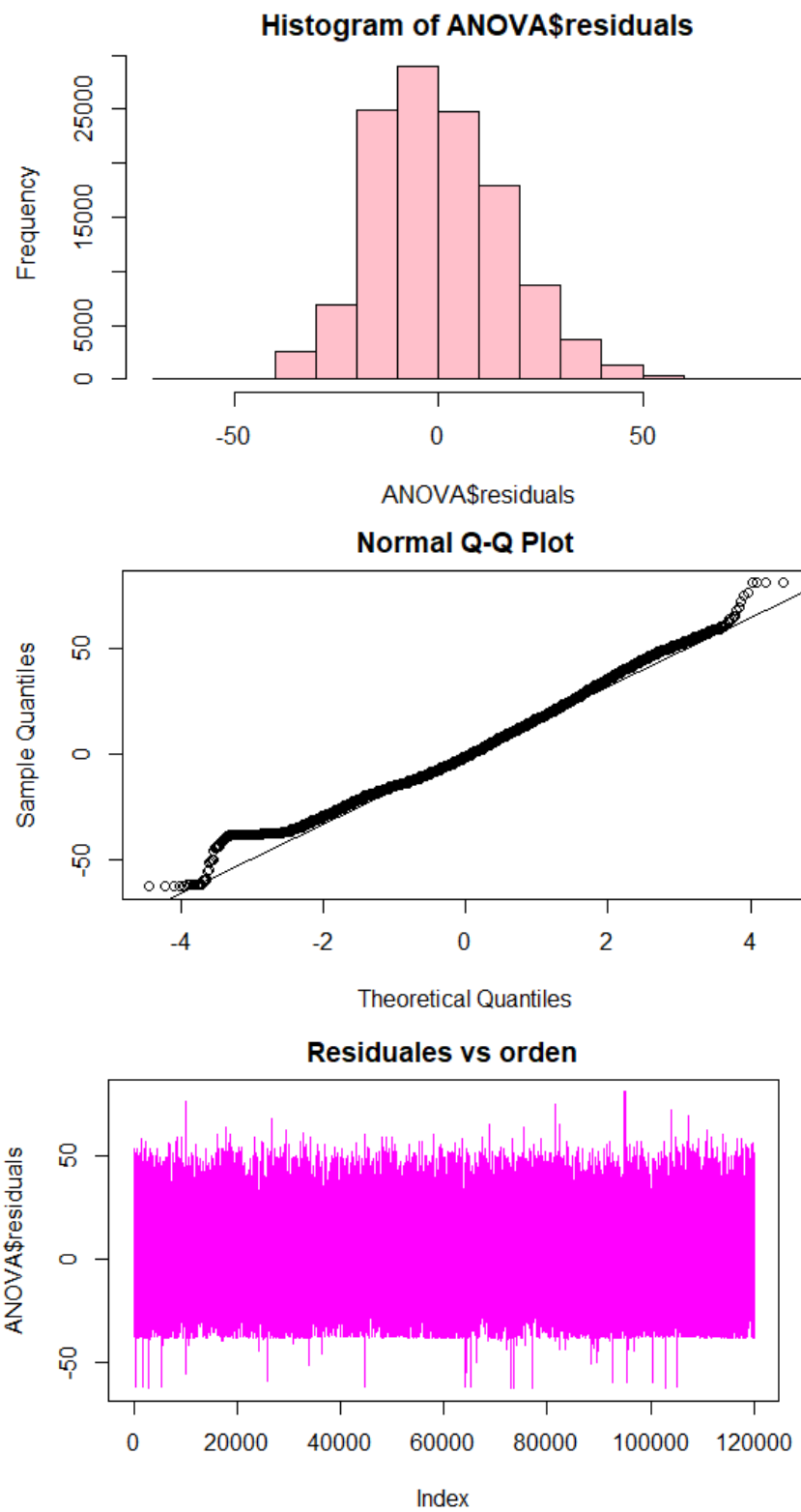
	Df	Sum Sq	Mean Sq	F value	Pr(>F)
tanova\$MUERTO	1	3147495	3147495	12124	<0.0000000000000002 ***

Residuals	119998	31152275	260		
-----------	--------	----------	-----	--	--

---

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

### Supuestos de normalidad



Como el modelo ANOVA pasó los supuestos de normalidad y el valor  $p$  es menos que  $\alpha$ , podemos rechazar la hipótesis nula; por ende, la edad no es independiente de la muerte. Es decir, la edad sí afecta la probabilidad de morir.



## Conclusión

Este análisis nos permitió resolver la pregunta detonadora. Para empezar, con el estudio de proporciones descubrimos que los hombres fallecen más que las mujeres. Esto no solo fue comprobado por este análisis, sino otras investigaciones han descubierto que el COVID-19 afecta de manera distinta y más grave a los hombres que a las mujeres.

Después, con el intervalo de confianza y las pruebas de hipótesis, descubrimos que el estado con la mayor cantidad de muertes no es el estado con mayor proporción de muertes. El estado con mayor proporción de muertes fue Michoacán.

Además, con las tablas bivariadas pudimos concluir que la obesidad, la intubación y la diabetes son variables que afectan/aumentan la mortalidad. Sabiendo esto, observamos que Michoacán es el estado con mayor proporción de personas con diabetes y el octavo con mayor proporción de intubados y personas con obesidad; lo cual puede explicar por qué es el estado con mayor proporción de muertes.

Finalmente, el análisis ANOVA nos permitió confirmar que la edad sí afecta en la mortalidad.

## Catálogo de Entidades

CLAVE_ENTIDAD	ENTIDAD_FEDERATIVA	ABREVIATURA
01	AGUASCALIENTES	AS
02	BAJA CALIFORNIA	BC
03	BAJA CALIFORNIA SUR	BS
04	CAMPECHE	CC
05	COAHUILA DE ZARAGOZA	CL
06	COLIMA	CM
07	CHIAPAS	CS
08	CHIHUAHUA	CH
09	CIUDAD DE MÉXICO	DF
10	DURANGO	DG
11	GUANAJUATO	GT
12	GUERRERO	GR
13	HIDALGO	HG
14	JALISCO	JC
15	MÉXICO	MC
16	MICHOACÁN DE OCAMPO	MN
17	MORELOS	MS
18	NAYARIT	NT
19	NUEVO LEÓN	NL
20	OAXACA	OC
21	PUEBLA	PL
22	QUERÉTARO	QT
23	QUINTANA ROO	QR
24	SAN LUIS POTOSÍ	SP
25	SINALOA	SL
26	SONORA	SR
27	TABASCO	TC
28	TAMAULIPAS	TS
29	TLAXCALA	TL
30	VERACRUZ DE IGNACIO DE LA LLAVE	VZ
31	YUCATÁN	YN
32	ZACATECAS	ZS
36	ESTADOS UNIDOS MEXICANOS	EUM

## Referencias

- Coronavirus Response.* (s. f.). Recuperado 6 de octubre de 2022, de <https://covid19.ciga.unam.mx/apps/covid-19-monitoreo-de-la-situacion-por-municipios/explore>
- Introducción a la COVID-19: métodos de detección, prevención, respuesta y control.* (s. f.). OpenWHO. Recuperado 5 de octubre de 2022, de [https://openwho.org/courses/introduccion-al-ncov#:~:text=Coronavirus%20es%20una%20gran%20familia,espiratorio%20agudo%20grave%20\(SARS\).](https://openwho.org/courses/introduccion-al-ncov#:~:text=Coronavirus%20es%20una%20gran%20familia,espiratorio%20agudo%20grave%20(SARS).)
- Instituto Tecnológico y de Estudios Superiores de Monterrey.* (s. f.). Reto. Canvas. Recuperado 5 de octubre de 2022, de <https://experiencia21.tec.mx/courses/323281/pages/reto>
- Nuevo coronavirus 2019.* (s. f.). Recuperado 5 de octubre de 2022, de <https://www.who.int/es/emergencias/diseases/novel-coronavirus-2019>
- Statista. (2022). *México: estados más poblados en 2020.* Recuperado 22 de octubre de 2022, de <https://es.statista.com/estadisticas/575948/numero-de-personas-en-mexico-por-entidad-federativa/>