

Factores biológicos en muertes por COVID-19 respecto a hombres y mujeres

Daniela Jiménez Téllez A01654798 Juan Carlos Corona Vega A01660135
Ricardo Camacho Castillo A01654132 Sebastián Espinosa López A01659912

Octubre, 2021

Contents

1	Introducción	1
2	Metodología	2
3	Análisis de Variables y Pruebas de Hipótesis	3
3.1	¿La proporción de hombres fallecidos de Europa es la misma de México?	5
3.2	¿La muerte por COVID-19 en personas con las morbilidades analizadas depende del sexo? . .	12
3.3	¿La muerte por COVID-19 depende del sexo?	13
4	Discusión y conclusiones	14
5	Referencias bibliográficas	14
6	Anexo	14

1 Introducción

El COVID-19, como cualquier enfermedad contagiosa, es un problema que afecta a todo tipo de personas sin importar su clase social, país de origen, etnia u cualquier otro factor. A pesar de que este virus no discrimina, se ha demostrado que algunas condiciones previas al contagio, como hipertensión, obesidad, EPOC, etc., producen un mayor riesgo en la mortalidad. De manera análoga, se ha observado en diversos lugares del mundo que la tasa de muertes por COVID-19 ha sido mayor en hombres que en mujeres a pesar de que la tasa de contagios sea prácticamente igual para cada uno. Un ejemplo es el caso de Europa Occidental, donde el número de muertes por Coronavirus ha sido del 69% en hombres. Igualmente, en EUA, casi el doble de hombres han muerto por lo mismo con respecto a las mujeres. (Henriques, M., 2020)

Las hipótesis de factores que influyen en esta diferencia en la mortalidad entre hombres y mujeres son tanto del tipo social (diferencias en puestos de trabajo, exposición, estilos de vida), como de tipo biológicos. Una teoría de Philip Goulder, profesor de inmunología en la universidad de Oxford, afirma que la respuesta inmunológica ante el virus es más fuerte en las mujeres. Esto

se debe a que la proteína encargada de detectar virus como el SARS-CoV-2 se encuentra en el cromosoma X. En caso de tener células inmunes, esta proteína se expresará al doble en las mujeres ya que tienen los cromosomas XX y solo una vez en los hombres, los cuales tienen solo un cromosoma X.

Datos recopilados por un grupo de estudiantes de Harvard (Feldsher, K., 2020) mostraron una gran diferencia entre las tasas de mortalidad de hombres y mujeres en distintos estados, teniendo en unos casi la misma proporción, mientras que en otros sucede el caso de tener el doble de mortalidad para los hombres. Estas evidencias sugieren que los factores de tipo social son los que más afectan. Sin embargo, son más difíciles de analizar ya que se presentan de muchos tipos y en diferentes contextos. Por otra parte, los datos sobre condiciones médicas (factores biológicos) sí son registrados y contables, por lo que su análisis es más sencillo.

Dicho lo anterior, el objetivo de este trabajo es identificar las diferencias entre las tasas de mortalidad por COVID-19 en hombres y mujeres. Se considerarán distintos factores biológicos como enfermedades crónicas y morbilidades, lo que permitirá contestar la siguiente pregunta: **¿cuál es la diferencia en la mortalidad por COVID-19 entre hombres y mujeres en México tomando en cuenta únicamente el factor biológico?**

2 Metodología

Para poder definir la pregunta rectora se tomaron en cuenta las diferentes variables que se proporcionan en la base de datos del Gobierno de México con respecto al COVID-19. De las 43 que hay, se usarán las siguientes:

- **Fecha de defunción:** se utilizará esta variable para asegurar que las personas fallecieron debido a COVID-19.
- **Sexo:** se utilizará esta variable para comparar las proporciones entre las defunciones por COVID-19.
- **Personas con diabetes:** se utilizará esta variable para analizar si la enfermedad influye en las muertes por COVID-19.
- **Personas con hipertensión:** se utilizará esta variable para analizar si la enfermedad influye en las muertes por COVID-19.
- **Personas con EPOC:** se utilizará esta variable para analizar si la enfermedad influye en las muertes por COVID-19.
- **Personas con enfermedad renal crónica:** se utilizará esta variable para analizar si la enfermedad influye en las muertes por COVID-19.
- **Personas con asma:** se utilizará esta variable para analizar si la enfermedad influye en las muertes por COVID-19.

Con base en estas, se hará un análisis exploratorio el cual permite analizar cada una de las variables y poder graficarlas para tener un mejor entendimiento de ellas.

NOTA: Debido a que el parámetro seleccionado en este trabajo son proporciones y diferencias de proporciones, al ser variables discretas, no es posible utilizar métodos de análisis de varianza (ANOVA para uno o dos factores). Sin embargo, existen otros métodos similares que van más allá del alcance de este curso.

2.0.1 Análisis exploratorio

```
## [1] "El número de variables es: "
```

```
## [1] 43
```

```
## [1] "Los nombres de las variables son: "
```

## [1]	"FECHA_ACTUALIZACION"	"ID_REGISTRO"	"ORIGEN"
## [4]	"SECTOR"	"ENTIDAD_UM"	"ENTIDAD_NAC"
## [7]	"ENTIDAD_RES"	"MUNICIPIO_RES"	"TIPO_PACIENTE"
## [10]	"FECHA_INGRESO"	"FECHA_SINTOMAS"	"FECHA_DEF"
## [13]	"INTUBADO"	"NEUMONIA"	"EDAD"
## [16]	"NACIONALIDAD"	"EMBARAZO"	"HABLA LENGUA_INDIG"
## [19]	"INDIGENA"	"DIABETES"	"EPOC"
## [22]	"ASMA"	"INMUSUPR"	"HIPERTENSION"
## [25]	"OTRA_COM"	"CARDIOVASCULAR"	"OBESIDAD"
## [28]	"RENAL_CRONICA"	"TABAQUISMO"	"OTRO_CASO"
## [31]	"TOMA_MUESTRA_LAB"	"RESULTADO_LAB"	"TOMA_MUESTRA_ANTIGENO"
## [34]	"RESULTADO_ANTIGENO"	"CLASIFICACION_FINAL"	"MIGRANTE"
## [37]	"PAIS_NACIONALIDAD"	"PAIS_ORIGEN"	"UCI"
## [40]	"SEXO"	"ID_unit"	"Prob"
## [43]	"Stratum"		

```
## [1] "El número de pacientes es: "
```

```
## [1] 100000
```

2.0.1.1 Reducción de los datos a casos positivos por COVID-19

De la muestra de 10,000 personas seleccionadas aleatoriamente se filtrarán los casos positivos por COVID-19 con el siguiente código:

2.0.1.2 Reducción de los datos positivos a muertes por COVID-19

Como se mencionó anteriormente, en este proyecto se analizará la manera en que los factores biológicos afectan las muertes por COVID-19 en hombres y mujeres. Por lo tanto, se reducirán los datos para tener solo las defunciones que han habido a causa de este virus ya que no es necesario tomar en cuenta a las personas que se recuperaron.

Una vez teniendo la reducción, se analizarán las siguientes variables:

3 Análisis de Variables y Pruebas de Hipótesis

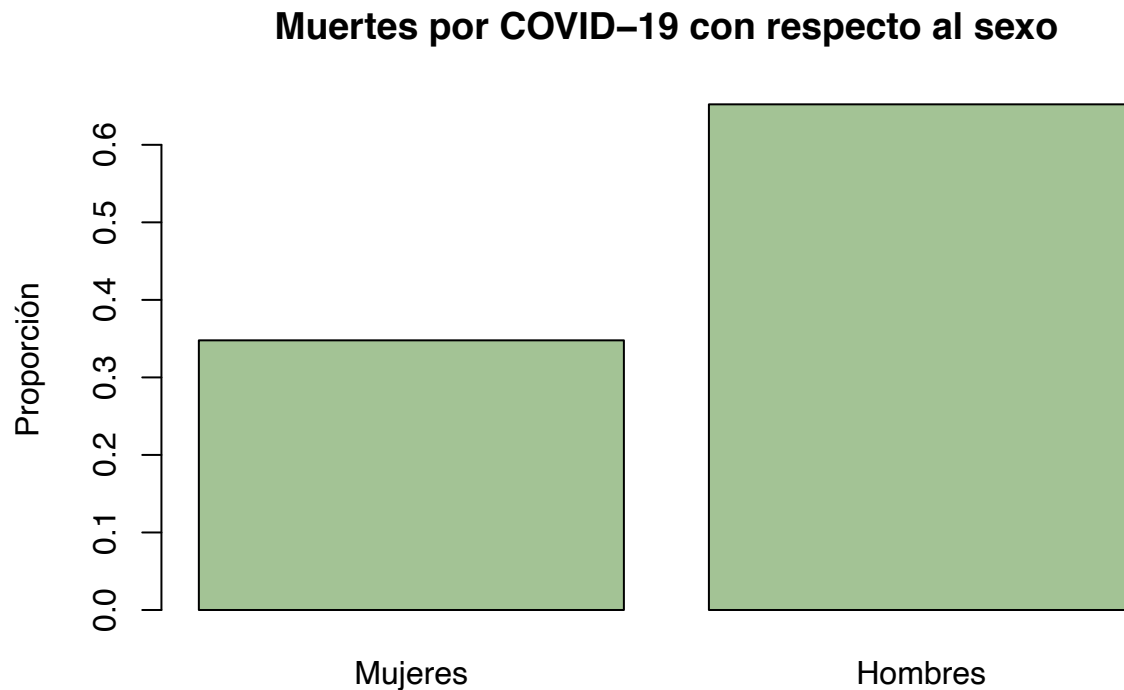
3.0.0.1 Análisis de la variable de sexo

```
## Mujeres Hombres
##      947      1776

## Total
##    2723

## Mujeres Hombres
##    0.3478 0.6522
```

```
## Diferencia
##      829
```



Al ver la tabla de proporciones entre hombres y mujeres, se puede notar que los hombres son más propensos a morir por COVID-19 que las mujeres. Para poder comprobar esto, se realizará una prueba de hipótesis de dos poblaciones (hombres y mujeres), la cual tiene como hipótesis nula (H_0) que las proporciones son iguales y como hipótesis alternativa (H_1) que los hombres mueren más.

$$H_0 : M - H = 0$$

$$H_1 : M - H < 0$$

Donde H es la proporción de hombres y M es la proporción de mujeres

```
##
## 2-sample test for equality of proportions without continuity
## correction
##
## data:  p out of n
## X-squared = 128.7, df = 1, p-value < 2.2e-16
## alternative hypothesis: less
## 95 percent confidence interval:
## -1.00000000 -0.02827458
```

```
## sample estimates:
##      prop 1      prop 2
## 0.06051505 0.09346876
```

Haciendo una prueba de hipótesis para la variable de sexo se puede decir que se rechaza la hipótesis nula ya que el valor de p es mucho menor que la significancia (0.05). Lo anterior confirma que las proporciones son diferentes, implicando que la proporción de hombres sea mayor ya que el valor de la diferencia de proporciones se encuentra dentro del intervalo de confianza calculado.

3.1 ¿La proporción de hombres fallecidos de Europa es la misma de México?

Según Henriques, M.(2020), en Europa Occidental el porcentaje de hombres que puere por COVID-19 es del 69%, por lo que creemos que en México esta proporción sera igual. Para acomprobar este supuesto planteamos la siguiente prueba de hipótesis:

$$H_0 : PH = 0.69$$

$$H_1 : PH \neq 0.69$$

```
## El intervalo de confianza de la proporción de hombres muertos por COVID-19
##      es: [ 0.6300496 , 0.6743504 ]
```

Al ver el intervalo $[0.6300496, 0.6743504]$, no hay evidencia para decir que la proporción real de hombres que fallecieron por COVID-19 en México es igual a la que hay en Europa Occidental, ya que el 0.69 de esta región no se encuentra dentro del intervalo de confianza calculado para México.

3.1.1 Análisis de la variable de diabetes con respecto al sexo

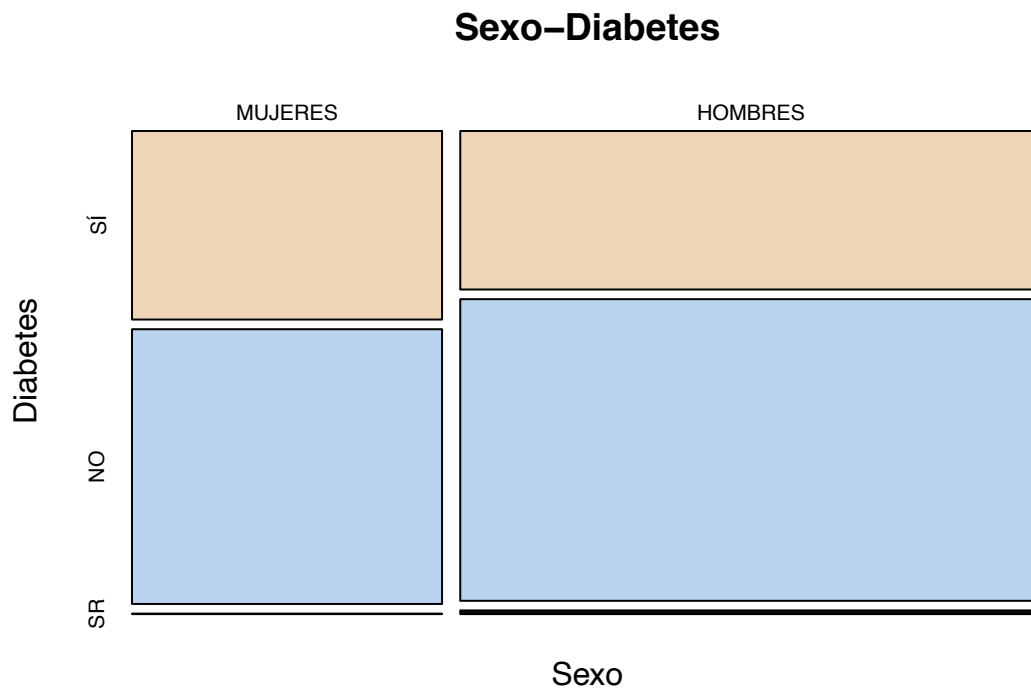
Para obtener un mejor entendimiento de la influencia de la morbilidad diabetes sobre el fallecimiento de ambos sexos, se calculan los intervalos de confianza para la diferencia de proporciones utilizando nivel de confianza del 95%

```
## [1] "La tabla en frecuencias:"

##           fdiabetes
## fsexo      SÍ    NO    SR
##  MUJERES  385  561    1
##  HOMBRES  607 1155   14

## [1] "La tabla en proporciones: "
```

	SÍ	NO	SR
MUJERES	0.1414	0.2060	0.0004
HOMBRES	0.2229	0.4242	0.0051



Viendo la tabla se sabe que hay 992 personas que tienen diabetes. De estas, 607 son hombres, lo que nos dice que son 8.15% más propensos a morir que las mujeres. Igualmente, se puede ver que hay más fallecidos por COVID-19 que no sufren de esta enfermedad.

```
## [1] 395.8704
```

```
## [1] 1380.13
```

```
## [1] 133.9058
```

```
## [1] 813.0942
```

```
## El intervalo de diferencia de proporciones de mujeres y hombres fallecidos
## con diabetes [ -0.09550088 , -0.06749912 ]
```

Del intervalo de confianza calculado podemos asegurar que la diferencia de proporciones es negativa, ya que la proporción de hombres que fallecieron y padecían de Diabetes es mayor a la de las mujeres. Esto confirma lo observado en la gráfica bidimensional.

3.1.2 Análisis de la variable de hipertensión con respecto al sexo

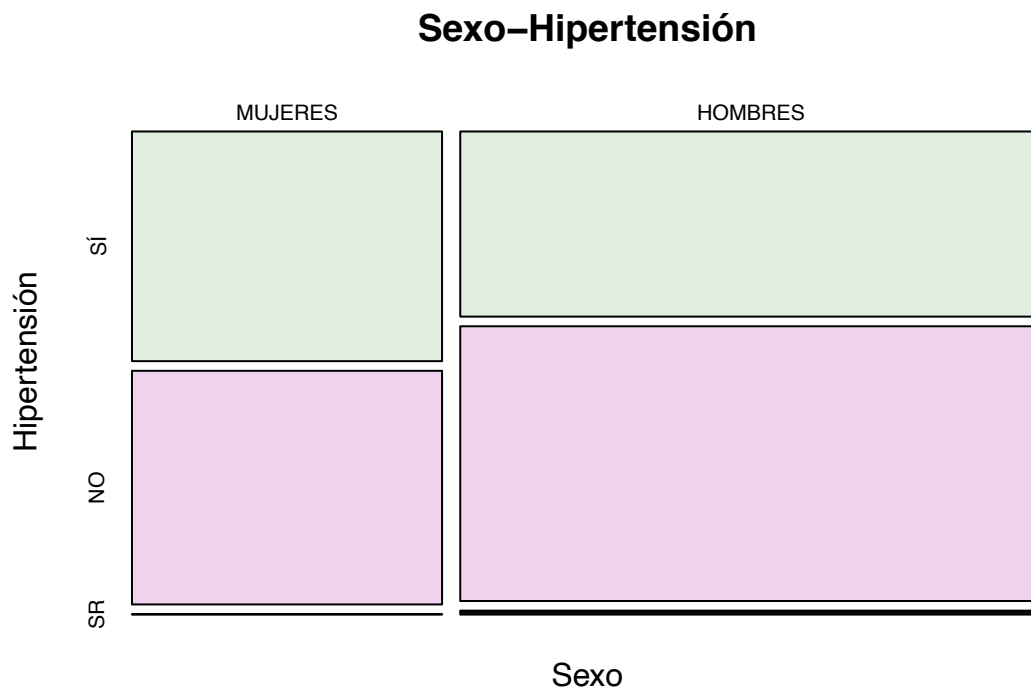
Para obtener un mejor entendimiento de la influencia de la morbilidad hipertensión sobre el fallecimiento de ambos sexos, se calculan los intervalos de confianza para la diferencia de proporciones utilizando nivel de confianza del 95%

```
## [1] "La tabla en frecuencias:"

##           fhipertension
## fsexo      SÍ    NO    SR
## MUJERES  469  477    1
## HOMBRES  709 1052   15

## [1] "La tabla en proporciones: "
```

	SÍ	NO	SR
MUJERES	0.1722	0.1752	0.0004
HOMBRES	0.2604	0.3863	0.0055



Por otro lado, al ver esta tabla se sabe que hay 1,178 personas que tienen hipertensión. De estas, 709 son hombres, lo que nos dice que son 8.82% más propensos a morir que las mujeres. Igualmente, se puede ver que hay más fallecidos por COVID-19 que no sufren de esta enfermedad.

```
## [1] 462.4704

## [1] 1313.53

## [1] 163.0734
```

```
## [1] 783.9266
```

```
## El intervalo de diferencia de proporciones de mujeres y hombres fallecidos  
## con hipertensión [ -0.1048998 , -0.07150015 ]
```

Del intervalo de confianza calculado podemos asegurar que la diferencia de proporciones es negativa, ya que la proporción de hombres que fallecieron y padecían de hipertensión es mayor a la de las mujeres. Esto confirma lo observado en la gráfica bidimensional.

3.1.3 Análisis de la variable de EPOC con respecto al sexo

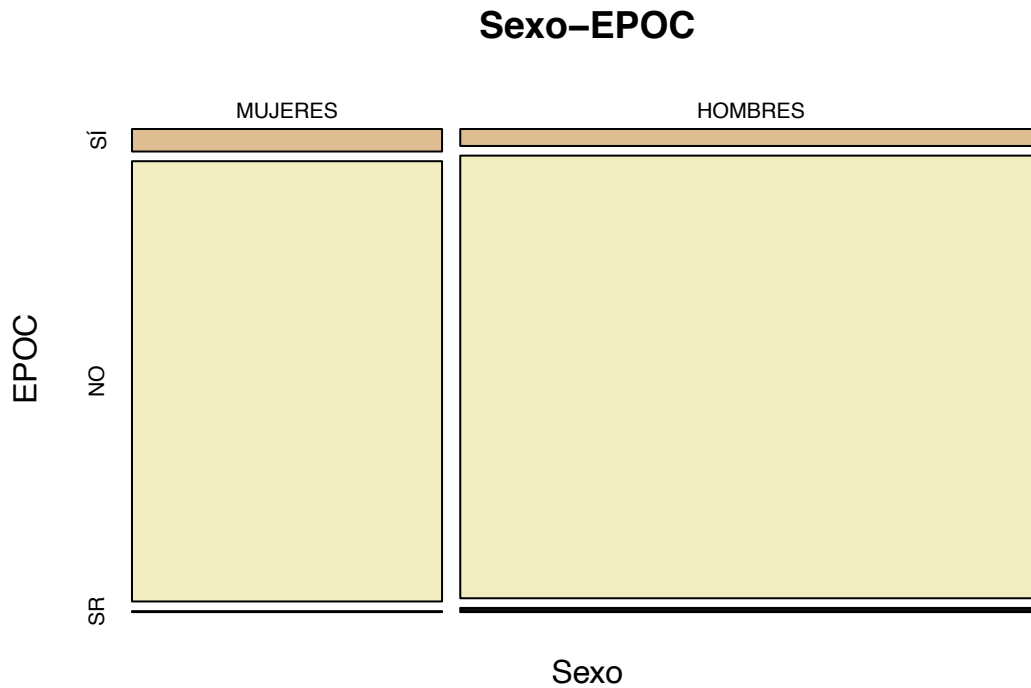
Para obtener un mejor entendimiento de la influencia de la morbilidad EPOC sobre el fallecimiento de ambos sexos, se calculan los intervalos de confianza para la diferencia de proporciones utilizando nivel de confianza del 95%

```
## [1] "La tabla en frecuencias:"
```

```
##          fepoc  
## fsexo      SÍ   NO   SR  
## MUJERES   46  899    2  
## HOMBRES   65 1695   16
```

```
## [1] "La tabla en proporciones: "
```

```
##          fepoc  
## fsexo      SÍ      NO      SR  
## MUJERES 0.0169 0.3302 0.0007  
## HOMBRES 0.0239 0.6225 0.0059
```

Para la tabla de EPOC se tiene que hay 111 personas que sufrían de esta enfermedad. De estas, 65 son hombres, lo que nos dice que son 0.7% más propensos a morir que las mujeres. Igualmente, se puede ver que hay más fallecidos por COVID-19 que no sufren de esta enfermedad.

```
## [1] 42.4464
```

```
## [1] 1733.554
```

```
## [1] 16.0043
```

```
## [1] 930.9957
```

```
## El intervalo de diferencia de proporciones de mujeres y hombres fallecidos
## con EPOC [ -0.008584958 , -0.005415042 ]
```

Del intervalo de confianza calculado podemos asegurar que la diferencia de proporciones es negativa, ya que la proporción de hombres que fallecieron y padecían de EPOC es mayor a la de las mujeres. Esto confirma lo observado en la gráfica bidimensional.

3.1.4 Análisis de la variable de enfermedad renal crónica con respecto al sexo

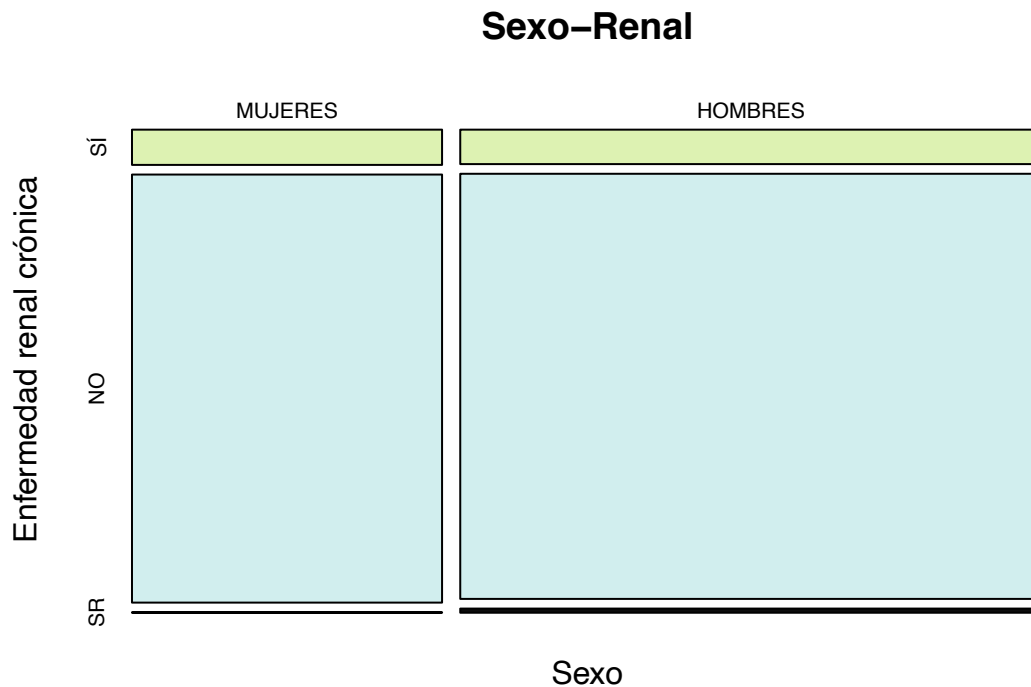
Para obtener un mejor entendimiento de la influencia de la morbilidad enfermedad renal crónica sobre el fallecimiento de ambos sexos, se calculan los intervalos de confianza para la diferencia de proporciones utilizando nivel de confianza del 95%

```
## [1] "La tabla en frecuencias:"

##           frenal
## fsexo      SÍ    NO    SR
## MUJERES    72  874    1
## HOMBRES   132 1628   16

## [1] "La tabla en proporciones: "

##           frenal
## fsexo      SÍ    NO    SR
## MUJERES 0.0264 0.3210 0.0004
## HOMBRES 0.0485 0.5979 0.0059
```



Para la tabla de enfermedad renal crónica se tiene que hay 204 personas que sufrían de esta enfermedad. De estas, 132 son hombres, lo que nos dice que son 2.21% más propensos a morir que las mujeres. Igualmente, se puede ver que hay más fallecidos por COVID-19 que no sufren de esta enfermedad.

```
## [1] 86.136

## [1] 1689.864

## [1] 25.0008
```

```
## [1] 921.9992
```

```
## El intervalo de diferencia de proporciones de mujeres y hombres fallecidos  
## con enfermedad renal [ -0.02492194 , -0.01927806 ]
```

Del intervalo de confianza calculado podemos asegurar que la diferencia de proporciones es negativa, ya que la proporción de hombres que fallecieron y padecían de una enfermedad renal crónica es mayor a la de las mujeres. Esto confirma lo observado en la gráfica bidimensional.

3.1.5 Análisis de la variable de asma con respecto al sexo

Para obtener un mejor entendimiento de la influencia de la morbilidad asma sobre el fallecimiento de ambos sexos, se calculan los intervalos de confianza para la diferencia de proporciones utilizando nivel de confianza del 95%

```
## [1] 19
```

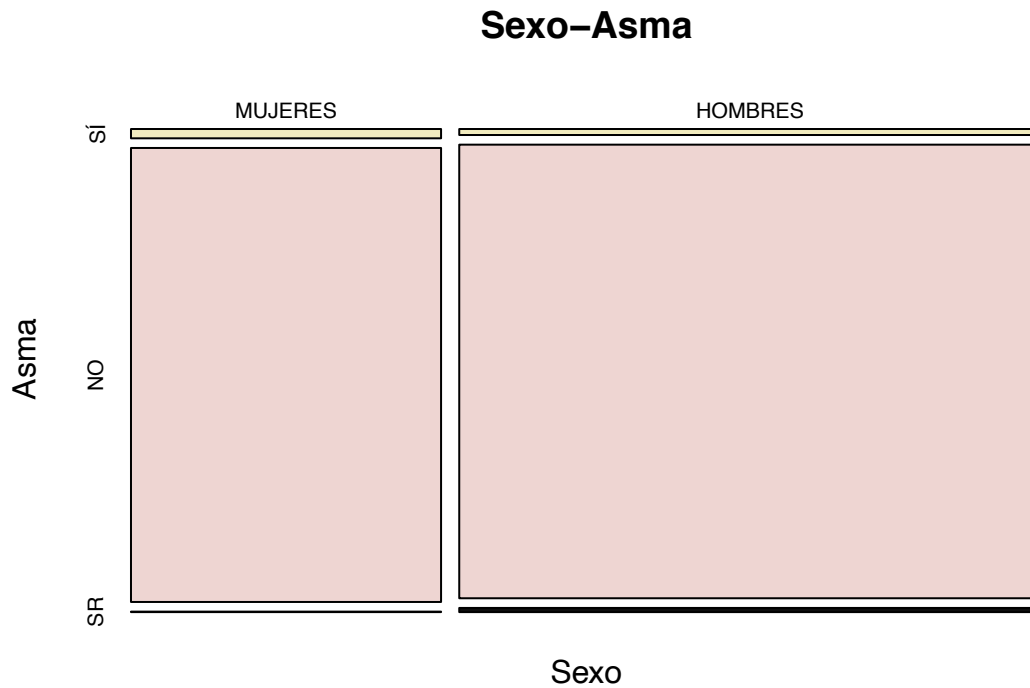
```
## [1] 23
```

```
## [1] "La tabla en frecuencias:"
```

```
##          fasma  
## fsexo      SÍ    NO    SR  
## MUJERES   19  927    1  
## HOMBRES   23 1737   16
```

```
## [1] "La tabla en proporciones: "
```

```
##          fasma  
## fsexo      SÍ      NO      SR  
## MUJERES 0.0070 0.3404 0.0004  
## HOMBRES 0.0084 0.6379 0.0059
```



Para la tabla de asma se tiene que hay 42 personas que sufrían de esta enfermedad. De estas, 23 son hombres, lo que nos dice que son 0.014% más propensos a morir que las mujeres. Igualmente, se puede ver que hay más fallecidos por COVID-19 que no sufren de esta enfermedad.

```
## [1] 14.9184
```

```
## [1] 1761.082
```

```
## [1] 6.629
```

```
## [1] 940.371
```

```
## El intervalo de diferencia de proporciones de mujeres y hombres fallecidos
## con enfermedad renal [ -0.002004734 , -0.0007952664 ]
```

Del intervalo de confianza calculado podemos asegurar que la diferencia de proporciones es negativa, ya que la proporción de hombres que fallecieron y padecían de Diabetes es mayor a la de las mujeres. Esto confirma lo observado en la gráfica bidimensional.

3.2 ¿La muerte por COVID-19 en personas con las morbilidades analizadas depende del sexo?

Para tener un mejor entendimiento de la relación de ambos factores se propone realizar una prueba de independencia (Chi cuadrada) para aceptar o rechazar la siguiente hipótesis:

H_0 : El fallecimiento de personas por COVID-19 con morbilidades es independiente de su sexo.

H_1 : El fallecimiento de personas por COVID-19 con morbilidades no es independiente de su sexo.

```
##
## Pearson's Chi-squared test
##
## data:  Q
## X-squared = 2.4306, df = 4, p-value = 0.6571

##      morbilidad
## sexo Diabetes Hipertensión      EPOC      Renal      Asma
## M 389.0273      461.9699 43.53027  80.00158 16.47091
## H 602.9727      716.0301 67.46973 123.99842 25.52909

##      morbilidad
## sexo  Diabetes Hipertensión      EPOC      Renal      Asma
## M -0.2041853      0.3270793 0.3743288 -0.8945953 0.6231674
## H 0.1640083      -0.2627207 -0.3006730 0.7185679 -0.5005482

## [1] 5.991465
```

Dado que valor- $p > 0.05$, se puede decir que las morbilidades dependen del sexo del paciente, esto quiere decir que los factores biológicos que seleccionados (Morbilidades) no representan información suficiente para comprobar nuestra hipótesis de que la causa de muerte por COVID se vea afectada por factores biológicos de cada sexo.

3.3 ¿La muerte por COVID-19 depende del sexo?

Para tener un mejor entendimiento de la relación de ambos factores se propone realizar una prueba de independencia (Chi cuadrada) para aceptar o rechazar la siguiente hipótesis:

H_0 : El fallecimiento de personas por COVID-19 es independiente de su sexo.

H_1 : El fallecimiento de personas por COVID-19 no es independiente de su sexo.

```
##
## Pearson's Chi-squared test with Yates' continuity correction
##
## data:  Qm
## X-squared = 109.87, df = 1, p-value < 2.2e-16

##      condicion
## sexo Muertos  Vivos
## M 1209.186 15386.81
## H 1513.814 19263.19

##      condicion
## sexo Muertos  Vivos
## M -7.539854 2.113660
## H 6.738655 -1.889059
```

[1] 5.991465

Al hacer la prueba de independencia de la relación entre la muerte y el género de los pacientes, se tienen los datos para decir que las variables no son independientes, por lo tanto se puede pensar que sí hay una relación entre el género y la probabilidad de fallecimiento por COVID-19.

4 Discusión y conclusiones

Una vez hecho el análisis de los datos proporcionados por el Gobierno de México, fue posible dar evidencia para afirmar que la proporción de hombres muertos por COVID-19 es mayor que la de las mujeres. Esto indica que hay un factor relacionado con el sexo el cual aumenta la probabilidad de fallecer a causa del virus; esto se confirma realizando una prueba de independencia entre factores.

Igualmente, al seguir la línea de investigación, se analizaron los datos de diferentes morbilidades y su presencia en pacientes fallecidos. Con esto, se obtuvo que la diferencia de proporción entre hombres y mujeres es distinta. Estas morbilidades no son dependientes del sexo y por lo tanto no se pueden tomar como un factor biológico que influye en la diferencia de proporciones de muerte entre hombres y mujeres.

Finalmente se concluyó que a pesar de que la probabilidad de muerte se ve influida por el sexo, los factores biológicos analizados no son los responsables de un aumento de probabilidad de muerte para algún sexo. De igual manera, los factores que generan esta diferencia son sociales u ocasionados por otras variables biológicas no consideradas en este análisis.

5 Referencias bibliográficas

Henriques, M. (12 de Abril de 2020). Why Covid-19 is different for men and woman. BBC future

Feldsher, K. (31 de Julio de 2020). More men than women are dying from Covid-19. Why?. Harvard T.H. Chan School of Public Health. <https://www.hsph.harvard.edu/news/features/more-men-than-women-are-dying-from-covid-19-why/>

Secretaría de salud. (19 de septiembre, 2021). Datos abiertos covid-19 [database]. <https://www.gob.mx/salud/documentos/datos-abiertos-152127>

6 Anexo

```

---
title: "Factores biológicos en muertes por COVID-19 respecto a hombres y mujeres"
author:
- Daniela Jiménez Téllez A01654798
- Juan Carlos Corona Vega A01660135
- Ricardo Camacho Castillo A01654132
- Sebastián Espinosa López A01659912
date: "Octubre, 2021"
output:
  pdf_document:
    latex_engine: xelatex
    toc: yes
    fig_caption: yes
    number_sections: yes
  html_document: default
---

```

```

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(knitr)
opts_chunk$set(tidy.opts=list(width.cutoff=40),tidy=TRUE)
```

```{r,echo=FALSE}

Carga de datos de la muestra
datos_muestra = read.csv("mcovidG5Eq5.csv")

Cantidad de variables
print("El número de variables es: ")
length(datos_muestra)

Nombre de las variables
print("Los nombres de las variables son: ")
names(datos_muestra)

Cantidad de pacientes
print("El número de pacientes es: ")
length(datos_muestra$ID_REGISTRO)

...

```{r,echo=FALSE}
# Reducción de los datos a casos positivos por COVID-19
positivos = datos_muestra[datos_muestra$CLASIFICACION_FINAL<4, ]
```

```{r,echo=FALSE}

```

```

fallecidos = positivos[positivos$FECHA_DEF!="9999-99-99", ]

# Guardar los datos en un Excel
write.csv(fallecidos, "mfallecidos.csv", row.names = FALSE, fileEncoding = "UTF-8")

...

```{r,echo=FALSE}
Reducción de los datos a casos positivos por COVID-19
positivos = datos_muestra[datos_muestra$CLASIFICACION_FINAL<4,]

...

```{r,echo=FALSE}

fallecidos = positivos[positivos$FECHA_DEF!="9999-99-99", ]

# Guardar los datos en un Excel
write.csv(fallecidos, "mfallecidos.csv", row.names = FALSE, fileEncoding = "UTF-8")

...

```{r,echo=FALSE}
Tabla de fallecidos con respecto al sexo
t_sexo = table(fallecidos$SEXO)
names(t_sexo) = c("Mujeres", "Hombres")
t_sexo

NH <- t_sexo["Hombres"]
NM <- t_sexo["Mujeres"]

nh <- as.numeric(NH)
nm <- as.numeric(NM)

total = t_sexo["Mujeres"]+t_sexo["Hombres"]
names(total) = ("Total")
total

Tabla en proporciones
tp_sexo = round(prop.table(t_sexo), 4)
tp_sexo

PM = tp_sexo["Mujeres"]
PH = tp_sexo["Hombres"]

dif_s = round(NH-NM, 4)

```



```

names(dif_s) = c("Diferencia")
dif_s
...

```{r,echo=FALSE}
# Gráfica de la tabla en proporciones
barplot(tp_sexo, ylab = "Proporción", col = "#A3C395",
        main = "Muertes por COVID-19 con respecto al sexo")
...

```{r,echo=FALSE}
Prueba de hipótesis e intervalos de diferencia
de proporciones para la variable de sexo
t_sexoPos <- table(positivos$SEXO)
names(t_sexoPos) <- c("Mujeres", "Hombres")
p <- c(as.numeric(t_sexo[1]),as.numeric(t_sexo[2]))
n <- c(as.numeric(t_sexoPos[1]),as.numeric(t_sexoPos[2]))
prop.test(p, n, conf.level = 0.95, correct = FALSE, alternative = "less")
...

```{r,echo=FALSE}
a = 0.05
z1 <- abs(qnorm(a/2))

limI <- PH - z1*sqrt(PH*(1-PH)/NH)
limS <- PH + z1*sqrt(PH*(1-PH)/NH)
cat("El intervalo de confianza de la proporción de hombres muertos por COVID-19
    es:", "[", limI, ", ", limS, "]\n")
...

```{r,echo=FALSE}

Tabla bivariada de fallecidos con respecto a las variables de diabetes y sexo
fsexo = fallecidos$SEXO
fdiabetes = fallecidos$DIABETES
tb_d = table(fsexo, fdiabetes)

Tabla bivariada con nombres de referencia
row.names(tb_d) = c("MUJERES", "HOMBRES")
colnames(tb_d) = c("Sí", "NO", "SR")
print("La tabla en frecuencias:")
tb_d

Tabla bivariada en porcentajes
print("La tabla en proporciones: ")
tp_d = round(prop.table(tb_d), 4)
tp_d

```

```

Gráfica bidimensional para la tabla bivariada
plot(tb_d, col = c("#EED5B7", "#B9D3EE", "#0D0D0D"),
 xlab = "Sexo", ylab = "Diabetes", main = "Sexo-Diabetes")
...

```{r PH Diabetes,echo=FALSE}
# Cálculo de intervalos de confianza con significancia del 0.05

ph = tp_d[2,1]

pm = tp_d[1,1]

a = 0.05

#Condiciones para usar la normal
nh*ph
nh*(1-ph)
nm*pm
nm*(1-pm)

# Todos son mayores 5, por tanto se puede usar la normal

# Regla de decisión
z <- abs(qnorm(0.05/2))

LimID = (pm - ph) - z*sqrt(ph*(pm)/nh + pm*(ph)/nm)
LimSD = (pm - ph) + z*sqrt(ph*(pm)/nh + pm*(ph)/nm)

cat("El intervalo de diferencia de proporciones de mujeres y hombres fallecidos
    con diabetes", "[", LimID, ", ", LimSD, "]\n")

...

```{r,echo=FALSE}

Tabla bivariada de fallecidos con respecto a las variables de hipertensión y sexo
fsexo = fallecidos$SEXO
fhipertension = fallecidos$HIPERTENSION
tb_h = table(fsexo, fhipertension)

Tabla bivariada con nombres de referencia
row.names(tb_h) = c("MUJERES", "HOMBRES")
colnames(tb_h) = c("SÍ", "NO", "SR")
print("La tabla en frecuencias:")
tb_h

Tabla bivariada en porcentajes

```

```

print("La tabla en proporciones: ")
tp_h = round(prop.table(tb_h), 4)
tp_h

Gráfica bidimensional para la tabla bivariada
plot(tb_h, col = c("#E0EEE0", "#EED2EE", "#0D0D0D"),
 xlab = "Sexo", ylab = "Hipertensión", main = "Sexo-Hipertensión")
...

```{r PH Hipertensión,echo=FALSE}
# Cálculo de intervalos de confianza con significancia del 0.05

ph = tp_h[2,1]

pm = tp_h[1,1]

a = 0.05

#Condiciones para usar la normal
nh*ph
nh*(1-ph)
nm*pm
nm*(1-pm)

# Todos son mayores 5, por tanto se puede usar la normal

# Regla de decisión
z <- abs(qnorm(0.05/2))

LimID = (pm - ph) - z*sqrt(ph*(pm)/nh + pm*(ph)/nm)
LimSD = (pm - ph) + z*sqrt(ph*(pm)/nh + pm*(ph)/nm)

cat("El intervalo de diferencia de proporciones de mujeres y hombres fallecidos
    con hipertensión", "[", LimID, ",", LimSD, "]\n")
...

```{r,echo=FALSE}

Tabla bivariada de fallecidos con respecto a las variables de EPOC y sexo
fsexo = fallecidos$SEXO
fepoc = fallecidos$EPOC
tb_e = table(fsexo, fepoc)

Tabla bivariada con nombres de referencia
row.names(tb_e) = c("MUJERES", "HOMBRES")
colnames(tb_e) = c("Sí", "NO", "SR")

```

```

print("La tabla en frecuencias:")
tb_e

Tabla bivariada en porcentajes
print("La tabla en proporciones: ")
tp_e = round(prop.table(tb_e), 4)
tp_e

Gráfica bidimensional para la tabla bivariada
plot(tb_e, col = c("#DEBD93", "#F2ECBF", "#0D0D0D"),
 xlab = "Sexo", ylab = "EPOC", main = "Sexo-EPOC")
...

```{r PH EPOC,echo=FALSE}
# Cálculo de intervalos de confianza con significancia del 0.05

ph = tp_e[2,1]

pm = tp_e[1,1]

a = 0.05

#Condiciones para usar la normal
nh*ph
nh*(1-ph)
nm*pm
nm*(1-pm)

# Todos son mayores 5, por tanto se puede usar la normal

# Regla de decisión
z <- abs(qnorm(0.05/2))

LimID = (pm - ph) - z*sqrt(ph*(pm)/nh + pm*(ph)/nm)
LimSD = (pm - ph) + z*sqrt(ph*(pm)/nh + pm*(ph)/nm)

cat("El intervalo de diferencia de proporciones de mujeres y hombres fallecidos
    con EPOC", "[", LimID, ", ", LimSD, "]\n")
...

```{r,echo=FALSE}

Tabla bivariada de fallecidos con respecto a las variables de enfermedad renal crónica y sexo
fsexo = fallecidos$SEXO
frenal = fallecidos$RENAL_CRONICA
tb_r = table(fsexo, frenal)

```

```

Tabla bivariada con nombres de referencia
row.names(tb_r) = c("MUJERES", "HOMBRES")
colnames(tb_r) = c("SÍ", "NO", "SR")
print("La tabla en frecuencias:")
tb_r

Tabla bivariada en porcentajes
print("La tabla en proporciones: ")
tp_r = round(prop.table(tb_r), 4)
tp_r

Gráfica bidimensional para la tabla bivariada
plot(tb_r, col = c("#DDF2B2", "#D1EEEE", "#OD0D0D"),
 xlab = "Sexo", ylab = "Enfermedad renal crónica", main = "Sexo-Renal")
...

```{r PH Renal,echo=FALSE}
# Cálculo de intervalos de confianza con significancia del 0.05

ph = tp_r[2,1]

pm = tp_r[1,1]

a = 0.05

#Condiciones para usar la normal
nh*ph
nh*(1-ph)
nm*pm
nm*(1-pm)

# Todos son mayores 5, por tanto se puede usar la normal

# Regla de decisión
z <- abs(qnorm(0.05/2))

LimID = (pm - ph) - z*sqrt(ph*(pm)/nh + pm*(ph)/nm)
LimSD = (pm - ph) + z*sqrt(ph*(pm)/nh + pm*(ph)/nm)

cat("El intervalo de diferencia de proporciones de mujeres y hombres fallecidos
    con enfermedad renal", "[", LimID, ", ", LimSD, "]\n")
...

```{r,echo=FALSE}

```

```

Tabla bivariada de fallecidos con respecto a las variables de asma y sexo
fsexo = fallecidos$SEXO
fasma = fallecidos$ASMA
tb_a = table(fsexo, fasma)
tb_a[1,1]
tb_a[2,1]

Tabla bivariada con nombres de referencia
row.names(tb_a) = c("MUJERES", "HOMBRES")
colnames(tb_a) = c("SÍ", "NO", "SR")
print("La tabla en frecuencias:")
tb_a

Tabla bivariada en porcentajes
print("La tabla en proporciones: ")
tp_a = round(prop.table(tb_a), 4)
tp_a

Gráfica bidimensional para la tabla bivariada
plot(tb_a, col = c("#F2ECBF", "#EED5D2", "#0D0D0D"),
 xlab = "Sexo", ylab = "Asma", main = "Sexo-Asma")

...

```{r PH Asma,echo=FALSE}
# Cálculo de intervalos de confianza con significancia del 0.05

ph = tp_a[2,1]

pm = tp_a[1,1]

a = 0.05

#Condiciones para usar la normal
nh*ph
nh*(1-ph)
nm*pm
nm*(1-pm)

# Todos son mayores 5, por tanto se puede usar la normal

# Regla de decisión
z <- abs(qnorm(0.05/2))

LimID = (pm - ph) - z*sqrt(ph*(pm)/nh + pm*(ph)/nm)
LimSD = (pm - ph) + z*sqrt(ph*(pm)/nh + pm*(ph)/nm)

```

```

cat("El intervalo de diferencia de proporciones de mujeres y hombres fallecidos
    con enfermedad renal", "[", LimID, ", ", LimSD, "]\n")

...

```{r,echo=FALSE}
d1=c(tb_d[1,1], tb_h[1,1], tb_e[1,1], tb_r[1,1], tb_a[1,1])
d2=c(tb_d[2,1], tb_h[2,1], tb_e[2,1], tb_r[2,1], tb_a[2,1])
Q<-as.table(rbind(d1,d2))
dimnames(Q)<-
list(sexo=c("M","H"),morbilidad=c("Diabetes","Hipertensión","EPOC","Renal","Asma"))
M<-chisq.test(Q)
M
M$expected
M$residuals
qchisq(0.95,2)

...

```{r,echo=FALSE}
D1=c(t_sexo["Mujeres"], t_sexoPos["Mujeres"])
D2=c(t_sexo["Hombres"], t_sexoPos["Hombres"])
Qm<-as.table(rbind(D1,D2))
dimnames(Qm)<-list(sexo=c("M","H"),condicion=c("Muertos", "Vivos"))
Mm<-chisq.test(Qm)
Mm
Mm$expected
Mm$residuals
qchisq(0.95,2)
...

```



Instituto Tecnológico y de Estudios Superiores de Monterrey

Campus Monterrey

Modelación estadística para la toma de decisiones

Dr. Armando Albert Huerta

Dra. Blanca Ruiz Hernández

Autoevaluación

Daniela Jiménez Téllez A01654798

Propósito

Hoy en día los datos son algo muy importante. El poder analizarlos nos da un mejor entendimiento de lo que está pasando y de lo que podría pasar. Durante este reto tuvimos que analizar la base de datos del COVID-19 que proporciona el Gobierno de México. El propósito de hacer esto es poder hacer uso de los métodos estadísticos que aprendimos en el curso y así poder llegar a una conclusión para la pregunta que planteamos. Igualmente, tener conocimiento de lo que se puede hacer con diferentes cantidades de datos y cómo podemos aplicarlo en el futuro.

Estrategia

Para completar el reto lo primero que hicimos fue analizar la base de datos de una manera general. A lo que me refiero con esto es que vimos la cantidad de pacientes que se habían registrado, los nombres de las variables que había en el excel, etcétera. Una vez hecho eso, filtramos los datos hasta poder tener la muestra de 10,000 personas con la que trabajamos. Teniendo nuestra muestra, planteamos una pregunta que nos interesara responder. Al principio no teníamos muy claro qué queríamos porque podíamos hacer muchas cosas interesantes con tantos datos y variables. Al final decidimos guiarnos por la pregunta: **¿Cuál es la diferencia en la mortalidad por COVID-19 entre hombres y mujeres tomando en cuenta únicamente el factor biológico en México?**. Elegimos las variables que creímos relevantes y analizamos cada una de ellas para sacar sus intervalos de confianza y valores p. Esto nos permitió llegar a una conclusión y así poder presentar nuestros resultados.

Siempre fue claro lo que teníamos que hacer; sin embargo, mientras más avanzamos con el curso, más métodos estadísticos podíamos aplicar en nuestro proyecto. Esto nos permitió tener un resultado más acertado y saber cómo interpretar los datos que nos arrojaba nuestro análisis de mejor manera.

Resultado

Habiendo dicho lo anterior, puedo decir que sí se logró el propósito de aprendizaje de este reto. A pesar de que hubieron algunos errores y cosas que se pueden mejorar a lo largo de todo este proyecto, aprendí muchas cosas que sé que aplicaré en mi futuro. Algunas de ellas fue que hay que tener un entendimiento de lo que se está haciendo y de lo que se pide. En estadística es muy importante entender el problema y la redacción de este. Igualmente, aprendí que los elementos que se ocupan, ya sean los datos, significancia, intervalos, etcétera, tienen que ser bien ocupados para tener un resultado coherente. Finalmente, una de las cosas más importantes que me llevo es que hay que saber interpretar los resultados. El saber qué significan los números que tenemos nos permite dar una buena conclusión a lo que estamos haciendo, igualmente, permite ayudar a las personas a las que se les presenta el proyecto a entender todo de una manera más clara.