

# Reporte Reto

José Abraham Martínez Licona\*      Juan Pablo Solís Ruiz†  
Mayra Stefany Gómez Triana‡      Annette Pamela Ruiz Abreu§

2022-10-11

## Contents

<b>Introducción</b>	<b>2</b>
<b>Objetivo</b>	<b>2</b>
<b>Pregunta Detonadora</b>	<b>2</b>
<b>Principales Variables de Interés</b>	<b>2</b>
<b>Análisis</b>	<b>2</b>
Lectura los datos . . . . .	2
Exploración de variables de interés . . . . .	2
Histograma de edades . . . . .	2
Proporción sexo . . . . .	4
Días hospitalizado hasta defunción . . . . .	6
Histograma pacientes en entidad médica . . . . .	6
Intubados . . . . .	7
Estudio de proporciones . . . . .	8
Creando una columna nueva . . . . .	8
Tabla de contingencia . . . . .	8
Edad promedio por estado . . . . .	9
Entidad médica y defunciones . . . . .	10
Top 5 entidades médicas con más y menos defunciones . . . . .	11
Intervalos de confianza . . . . .	13
Intervalo de confianza de defunción en los estados con más defunciones . . . . .	13
Conclusión . . . . .	14
Intervalo de confianza de defunción en los estados con más defunciones . . . . .	14
Conclusión . . . . .	15
Intervalo de confianza de la varianza de las edades con respecto de la muestra con la población . . . . .	16
Conclusión . . . . .	16
<b>Referencias</b>	<b>16</b>

---

\*ITESM, A01368551@tec.mx

†ITESM, A01067387@tec.mx

‡ITESM, A01625609@tec.mx

§ITESM, A0142359@tec.mx

## 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.

## 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? ¿Este resultado se ha mantenido similar al de hace un año? ¿La edad y la entubación son factores para la defunción de los pacientes?

## Principales Variables de Interés

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

## Análisis

### Lectura los datos

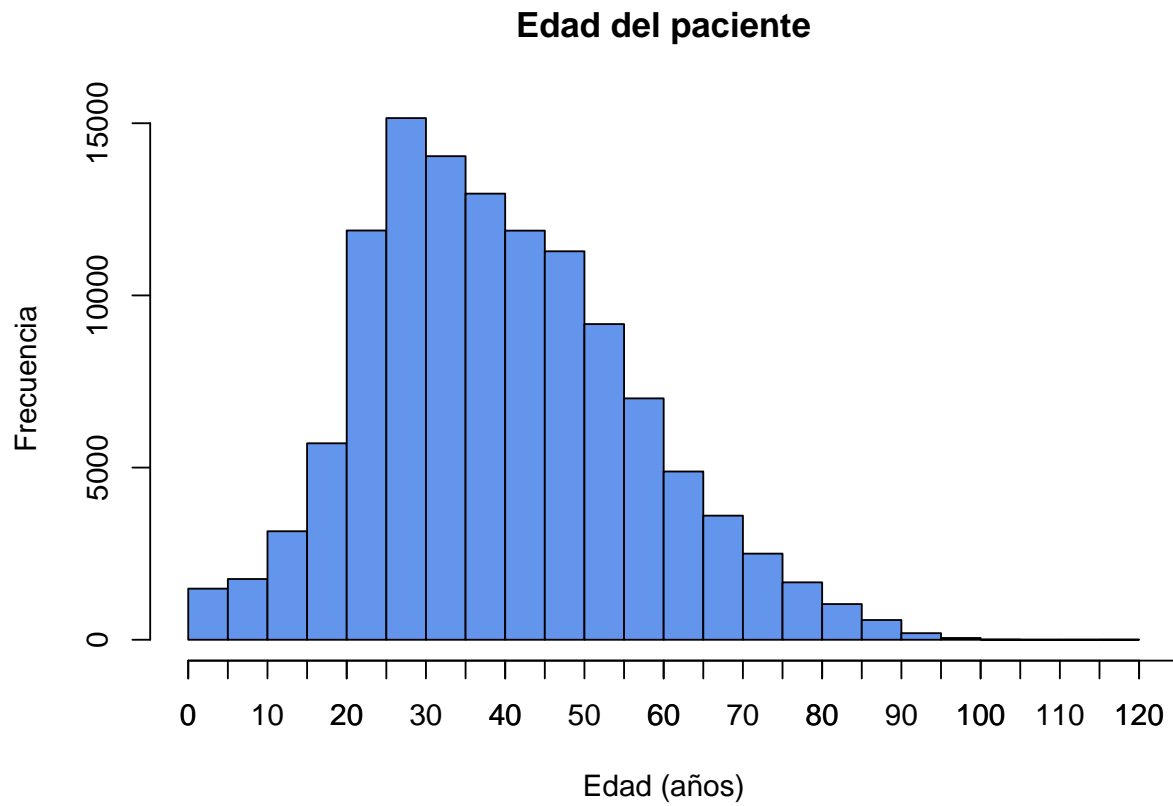
```
M = read.csv('Muestra120k_Eq8.csv',encoding = 'UTF-8')
```

### Exploración de variables de interés

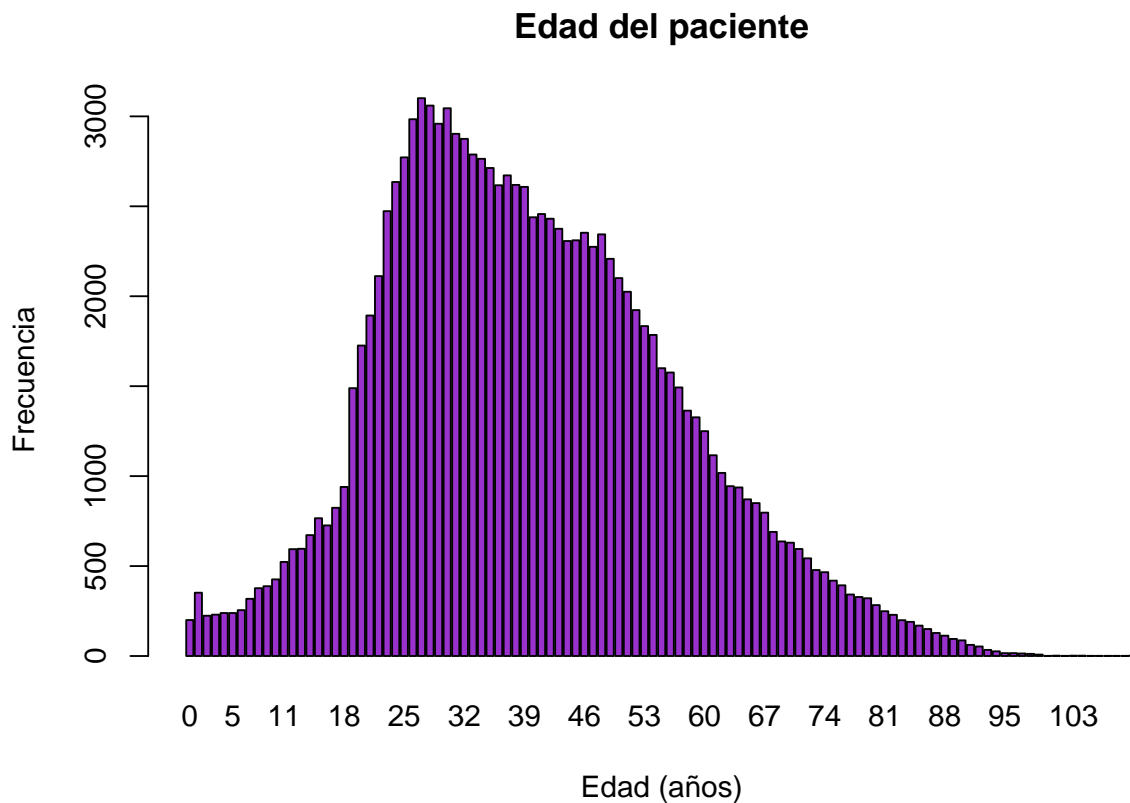
#### Histograma de edades

```
hist(M$EDAD, main = 'Edad del paciente',xlab = 'Edad (años)',  
ylab = 'Frecuencia',col = 'cornflowerblue')
```

```
axis(side=1, at=seq(0,150, 5))
```



```
T = table(M$EDAD)
barplot(T,main = 'Edad del paciente',xlab = 'Edad (años)',
ylab = 'Frecuencia',col = 'darkorchid')
```



#### Proporción sexo

```

T2 = table(M$SEX0)
names(T2) = c('Mujer', 'Hombre')
T3 = prop.table(T2)
T3

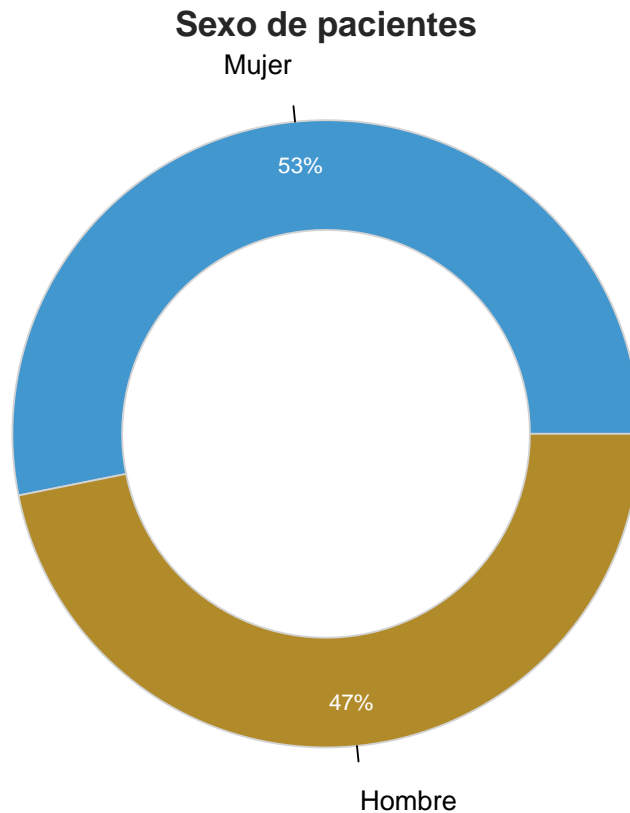
##  Mujer Hombre
## 0.5315 0.4685

library('lessR')

##
## lessR 4.2.3                                     feedback: gerbing@pdx.edu
## -----
## > d <- Read("")   Read text, Excel, SPSS, SAS, or R data file
##   d is default data frame, data= in analysis routines optional
##
## Learn about reading, writing, and manipulating data, graphics,
## testing means and proportions, regression, factor analysis,
## customization, and descriptive statistics from pivot tables.
##   Enter:  browseVignettes("lessR")
##
## View changes in this and recent versions of lessR.
##   Enter: news(package="lessR")
##
## **New Feature**: Interactive analysis of your data

```

```
## Enter: interact()
PieChart(T2,main = 'Sexo de pacientes', values="%",
          suggest=getOption("FALSE")) # lollipop plot
## >>> Note: T2 is not in a data frame (table)
## >>> Note: T2 is not in a data frame (table)
```

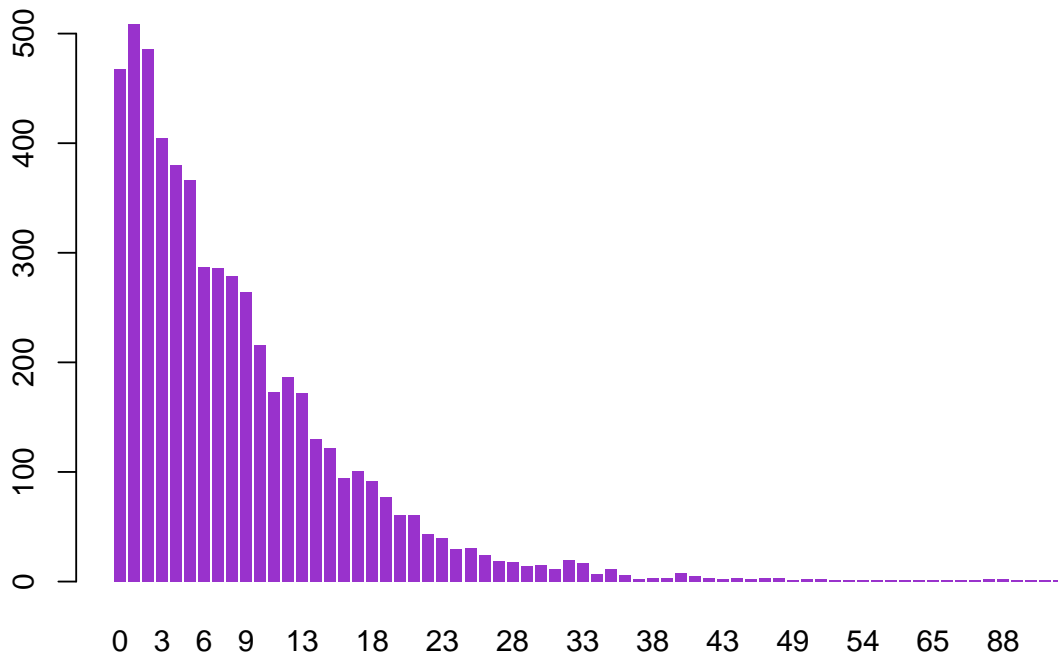


```
## >>> suggestions
## piechart(T2, hole=0) # traditional pie chart
## piechart(T2, values="%") # display %'s on the chart
## piechart(T2) # bar chart
## plot(T2) # bubble plot
## plot(T2, values="count") # lollipop plot
##
## --- T2 ---
##
##           Mujer  Hombre      Total
## Frequencies:  63780   56220   120000
## Proportions:   0.531   0.469     1.000
##
## Chi-squared test of null hypothesis of equal probabilities
##   Chisq = 476.280, df = 1, p-value = 0.000
```

## Días hospitalizado hasta defunción

### Limpieza

```
Mf = subset(M, M$FECHA_DEF != '9999-99-99') # and & or | not !+ ==
diash = as.numeric(as.Date(Mf$FECHA_DEF) - as.Date(Mf$FECHA_INGRESO))
diash1 = subset(diash, diash >= 1)
barplot(table(diash1), col = 'darkorchid', border = FALSE)
```

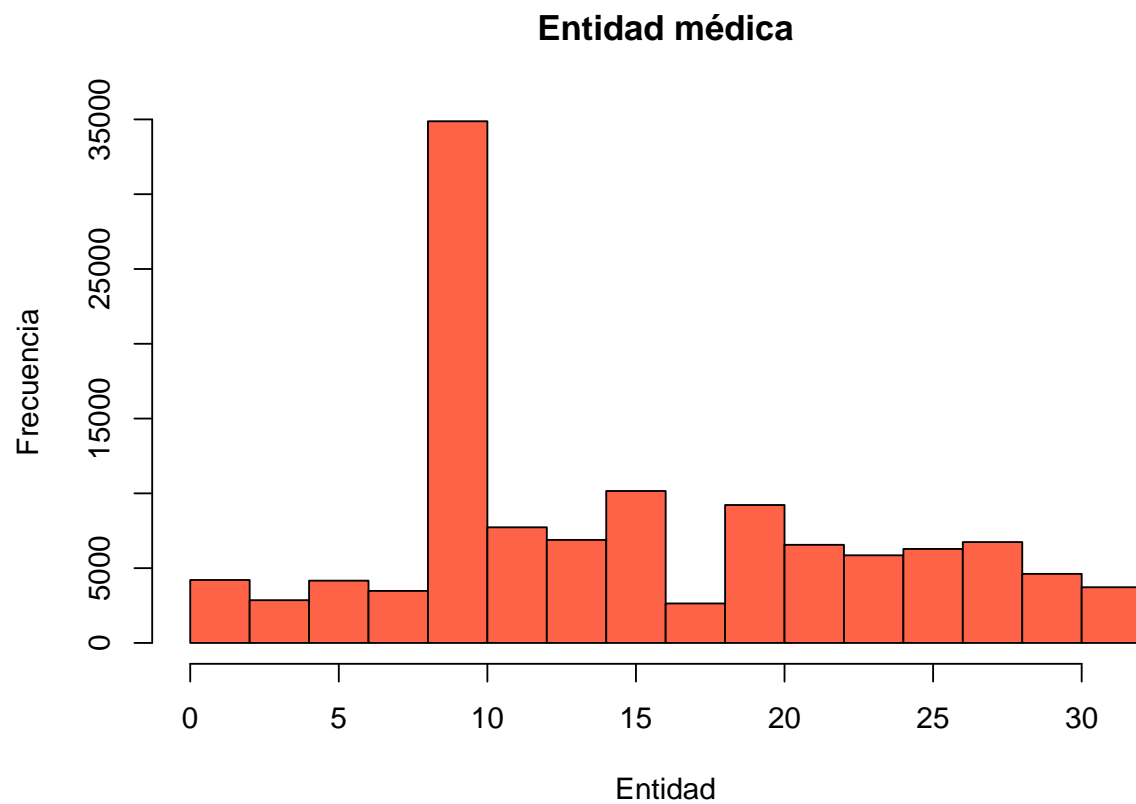


### Histograma pacientes en entidad médica

```
names = c("AS", "BC", "BS", "CC", "CL", "CM",
           "CS", "CH", "DF", "DG", "GT", "GR",
           "HG", "JC", "MC", "MN", "MS", "NT",
           "NL", "OC", "PL", "QT", "QR", "SP",
           "SL", "SR", "TC", "TS", "TL", "VZ",
           "YN", "ZS", "EUM")

a = c(1:32)

for(i in 1:32){
  ABREVIATURAS <- ifelse(M$ENTIDAD_UM == i, names[i], M$ENTIDAD_UM)
}
M = cbind(M, ABREVIATURAS)
hist(M$ENTIDAD_UM, main = 'Entidad médica', xlab = 'Entidad',
     ylab = 'Frecuencia', col = 'tomato')
```



#### Intubados

```

T = table(M$INTUBADO)
T1 = T[-3:-4]
names(T1) = c("SÍ", "NO")

T2 = round(prop.table(T1), 3)
cat("  INTUBADO\n")

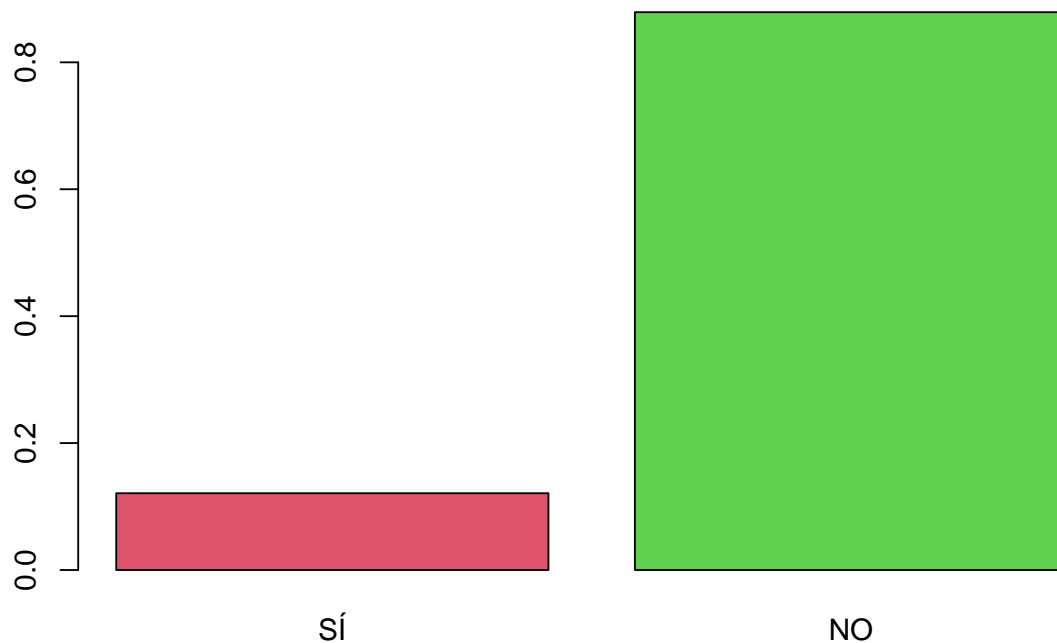
##    INTUBADO

T2

##    SÍ    NO
## 0.121 0.879

barplot(T2, col=2:3)

```



## Estudio de proporciones

### Creando una columna nueva

Para facilitar el análisis, se creó una columna nueva en donde 0 significa que el paciente murió y 1 significa que no murió.

```
MUERTO <- ifelse(M$FECHA_DEF == "9999-99-99",
                  0, 1)
M <- cbind(M, MUERTO)
```

### Tabla de contingencia

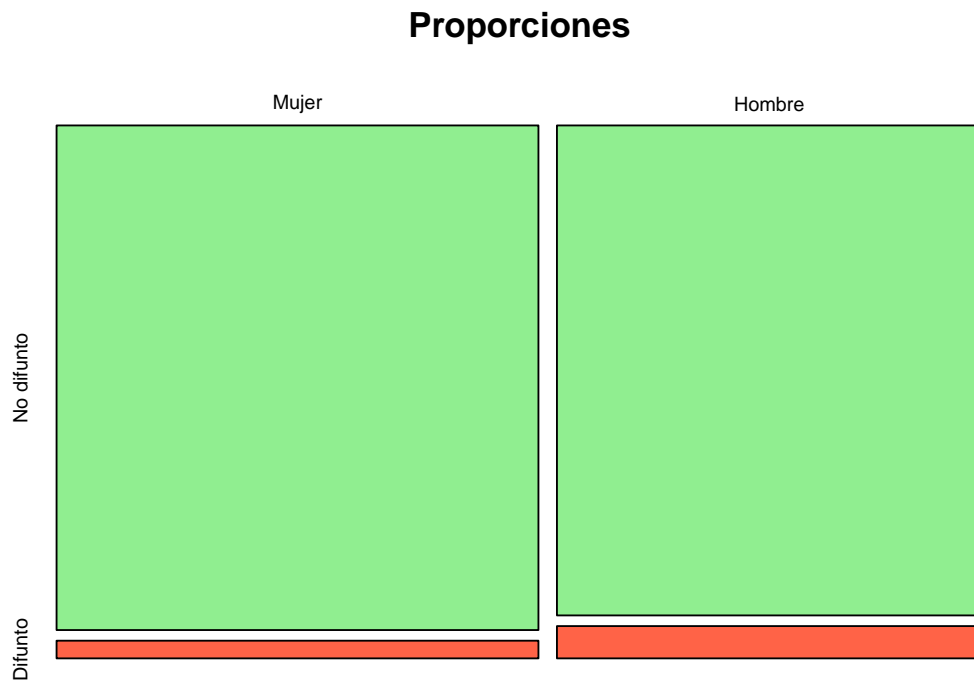
```
Ts = table(M$SEXO, M$MUERTO)
colnames(Ts) = c('No difunto', 'Difunto')
rownames(Ts) = c('Mujer', 'Hombre')
cat('      Defunción')

##      Defunción
##
Ts2 = round(prop.table(Ts), 3)
Ts2

##
##      No difunto Difunto
##  Mujer      0.514  0.018
##  Hombre      0.440  0.029
```



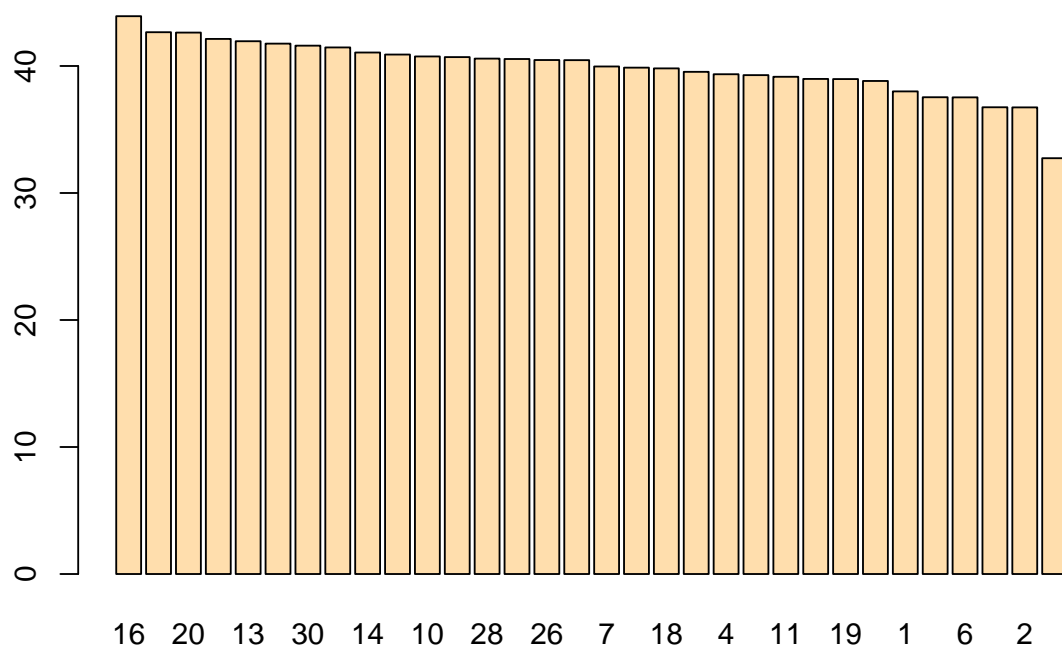
```
plot(Ts2, col = c('palegreen2','tomato'), main = 'Proporciones',c('1','2'))
```



### Edad promedio por estado

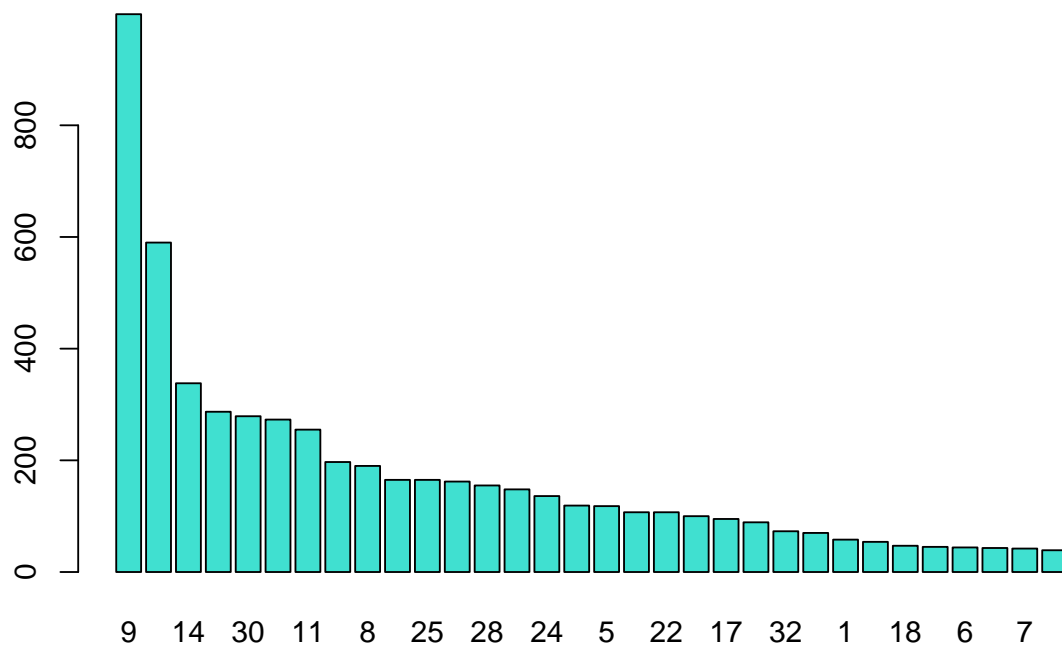
```
Td = tapply(M$EDAD, M$ENTIDAD_NAC, mean)
# Podemos cambiar mean por sd, length, mode, median, var, etc...
Td2 = round(Td, 2)[-33]
Td3 = sort(Td2, decreasing=TRUE)
barplot(Td3, col='navajowhite', main="Edad promedio por estado")
```

### Edad promedio por estado



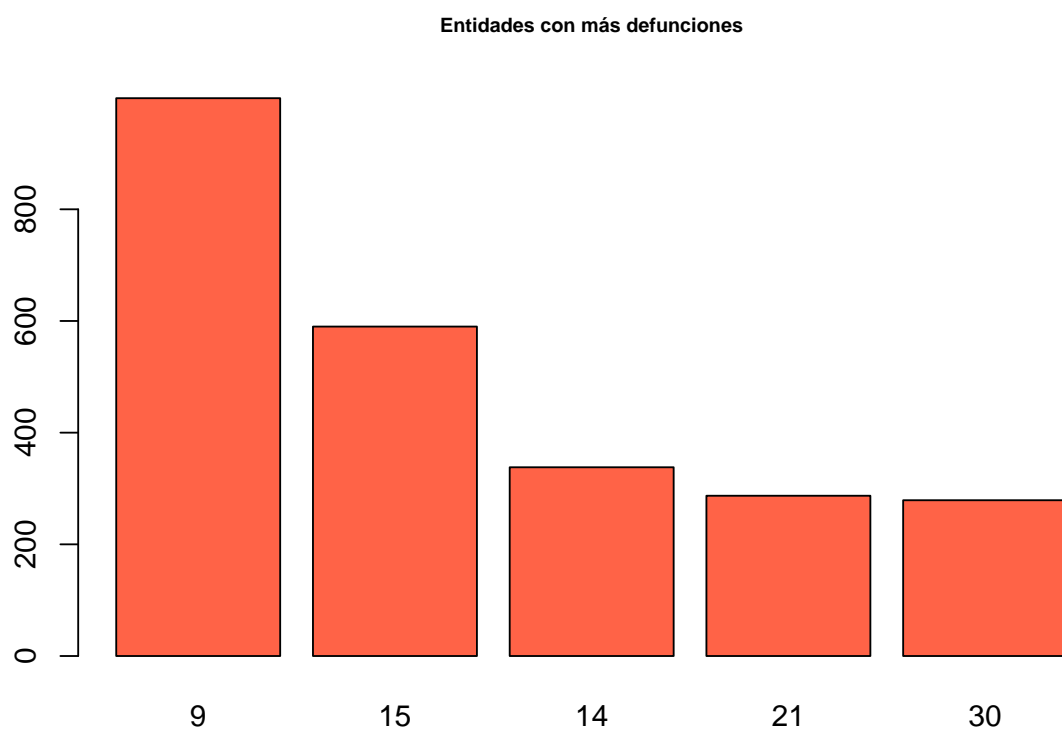
### Entidad médica y defunciones

```
Td = tapply(M$MUERTO, M$ENTIDAD_UM, sum)
# sd, length, var, ...
Td2 = round(Td, 2)
Td3 = sort(Td2, decreasing = TRUE)
barplot(Td3, col = 'turquoise')
```

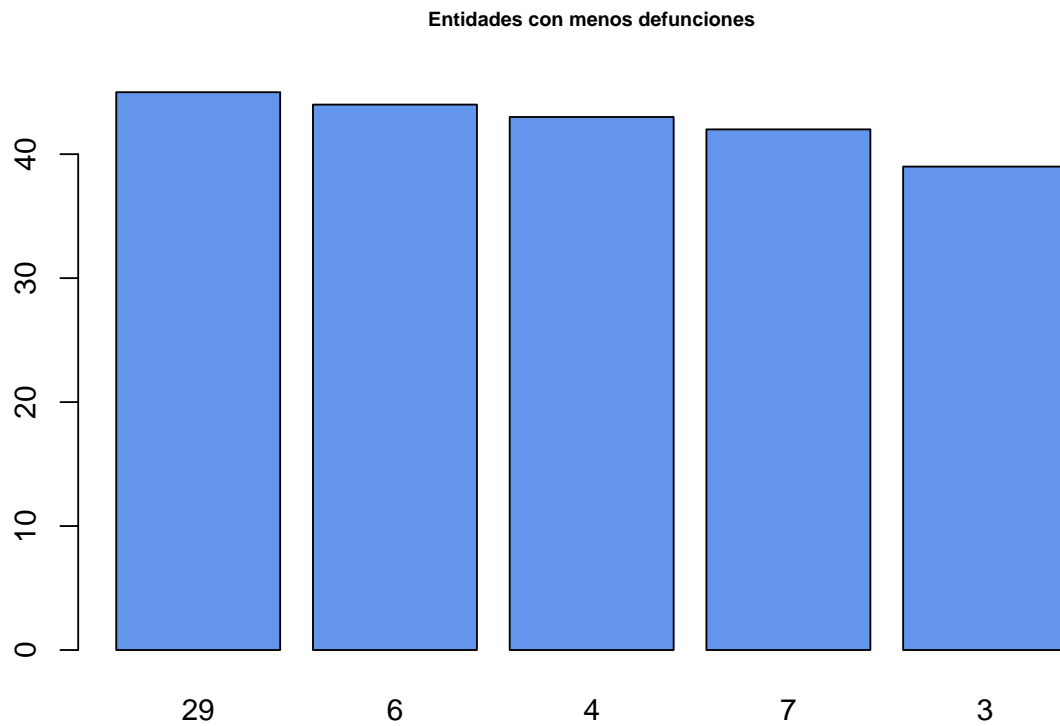


### Top 5 entidades médicas con más y menos defunciones

```
barplot(Td3[c(1:5)], col = 'tomato', main = "Entidades con más defunciones", cex.main=0.7)
```



```
barplot(Td3[c(28:32)], col = 'cornflowerblue', main = "Entidades con menos defunciones", cex.main=0.7)
```



## Intervalos de confianza

### Intervalo de confianza de defunción en 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.

$H_0$  :  $p_{DF} = p_{MC}$

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

$H_1$  :  $p_{DF} > p_{MC}$

```
DF = subset(M, M$ENTIDAD_UM==9)
```

```
MC = subset(M, M$ENTIDAD_UM==15)
```

```
cat("  MUERTOS\n")
```

```
##    MUERTOS
```

```
T_DF = table(DF$MUERTO)
```

```
cat("DF\n")
```

```
## DF
```

```
names(T_DF) = c("NO", "SÍ")
```

```
T_DF
```

```
##    NO    SÍ
```

```
## 32586    999

T_MC = table(MC$MUERTO)
cat("MC\n")

## MC

names(T_MC) = c("NO", "SÍ")
T_MC

##    NO    SÍ
## 7760   590

f1 = T_DF[2]
f2 = T_MC[2]
n1 = sum(T_DF)
n2 = sum(T_MC)
p = c(f1,f2)
n = c(n1,n2)
NC = 0.99
ICpp = prop.test(p, n, conf.level=NC, correct=FALSE)
ICpp

##
## 2-sample test for equality of proportions without continuity correction
##
## data:  p out of n
## X-squared = 307.05, df = 1, p-value < 0.00000000000000022
## alternative hypothesis: two.sided
## 99 percent confidence interval:
##  -0.04852113 -0.03330539
## sample estimates:
##      prop 1      prop 2
## 0.02974542 0.07065868

ICpp$conf.int

## [1] -0.04852113 -0.03330539
## attr(,"conf.level")
## [1] 0.99
```

## Conclusión

Como el 0 no está dentro del intervalo de confianza de la diferencia de proporciones entre DF y MC es de  $[-0.0485, -0.0333]$ , se estima, con un nivel de confianza del 99 %, que las proporciones en estos estados son diferentes; por ende, se rechaza la hipótesis nula. Como el intervalo es negativo, se estima que la proporción de muertos en el Estado de México es mayor que la del Distrito Federal.

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

$H_0$  : La proporción de muertos en Baja California Sur es igual a la proporción de muertos en Chiapas.

$H_0$  :  $p_{BS} = p_{CS}$

$H_1$  : La proporción de muertos en Baja California Sur es menor que la proporción de muertos en Chiapas.

$H_1$  :  $p_{BS} < p_{CS}$

CS = subset(M, M\$ENTIDAD\_UM==7)

BS = subset(M, M\$ENTIDAD\_UM==3)

```

cat("    MUERTOS\n\n")
##    MUERTOS
cat("CS    \n")
## CS

T_CS = table(CS$MUERTO)
names(T_CS) = c("NO", "SÍ")
T_CS

## NO  SÍ
## 730 42

cat("BS    \n")
## BS

T_BS = table(BS$MUERTO)
names(T_BS) = c("NO", "SÍ")
T_BS

## NO  SÍ
## 2061 39

f1 = T_CS[2]
f2 = T_BS[2]
n1 = sum(T_CS)
n2 = sum(T_BS)
p = c(f1,f2)
n = c(n1,n2)
NC = 0.99
ICpp = prop.test(p, n, conf.level=NC, correct=FALSE)
ICpp

##
## 2-sample test for equality of proportions without continuity correction
##
## data:  p out of n
## X-squared = 26.445, df = 1, p-value = 0.0000002712
## alternative hypothesis: two.sided
## 99 percent confidence interval:
##  0.01347828 0.05818715
## sample estimates:
##      prop 1      prop 2
## 0.05440415 0.01857143

ICpp$conf.int

## [1] 0.01347828 0.05818715
## attr(,"conf.level")
## [1] 0.99

```

## Conclusión

Como el 0 no está dentro del intervalo de confianza de la diferencia de proporciones entre CS y BS es de [0.0133, 0.058], se estima, con un nivel de confianza del 99 %, que las proporciones en estos estados son diferentes; por ende, se rechaza la hipótesis nula. Como el intervalo es positivo, se estima que la proporción de muertos en el Chiapas es mayor que la de Baja California Sur.

Intervalo de confianza de la varianza de las edades con respecto de la muestra con la población

$$H_0 : s^2 = \sigma^2$$

$$H_1 : s^2 \neq \sigma^2$$

```
ED=(M$EDAD)
```

```
n=length(ED)
var= var(ED)
alfa1=(1-0.95)/2
alfa2=(1-alfa1)
x21=qchisq(alfa1,n-1)
x22=qchisq(alfa2,n-1)
sup=((n-1)*var)/x21
inf=((n-1)*var)/x22

cat("[",inf,":",sup,"] al 0.95\n")
## [ 283.5604 : 288.1348 ] al 0.95
```

### Conclusión

Se considera que la varianza de edad de la muestra es mayor a la poblacional, ya que el intervalo de confianza obtenido no incluye el 1.

## Referencias

*Introducción a la COVID-19: métodos de detección, prevención, respuesta y control.* (s. f.). Open-WHO. Recuperado 5 de octubre de 2022, de [https://openwho.org/courses/introduccion-al-ncov#:%7E:text=Coronavirus%20es%20una%20gran%20familia,espiratorio%20agudo%20grave%20\(SARS\).](https://openwho.org/courses/introduccion-al-ncov#:%7E: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/emergencies/diseases/novel-coronavirus-2019>

*Coronavirus Response.* (s. f.). Recuperado 6 de octubre de 2022, de <https://covid19.ciga.unam.mx/apps/covid-19-monitoreo-de-la-situaci%C3%B3n-por-municipios/explore>