

Infografia Final - Probabilidad y Estadística Grupo B

Rubén Darío Vargas Yandy

6/15/2021

Introducción

Para la realización de las pruebas, requeridas para respaldar la infografía adjunta a este documento se realizó un preprocesamiento de los datos, para separarlos por medidas de tiempo, dada la facilidad que este tipo de operaciones tiene en lenguajes como Python3, haciendo uso de las librerías para el manejo de conjuntos de datos Pandas, Scikit-Learn y algunas built-in del mismo lenguaje.

El resumen de dicho pre-procesamiento se puede ver directamente en el archivo .ipynb de este directorio, o bien se puede leer su impresión y resumen en pdf, que se encuentra dentro del archivo “pre-procesamiento.pdf”

Bases de datos usadas:

- Casos Positivos de Covid-19 Colombia (fuente: Datos Abiertos Colombia)
- Reporte diario sobre Covid-19 Ministerio de la Salud de Colombia (fuente: cuenta de Twitter del Ministerio de la Salud de Colombia @MinSaludCol) recolectados por un bot, disponible su código en “scraper.py”

Pre-procesamiento:

Debido a la facilidad de filtrar conjuntos de datos por tiempo en Python3 usando la librería para Dataframes **Pandas** se realizó una separación de los conjuntos de datos en subconjuntos según la hipótesis a validar y se guardaron en la carpeta “out” de este directorio, los archivos en la carpeta “out” son los usados directamente en este archivo R Markdown.

Afirmación 1: Los grupos vulnerables siguen siendo los adultos mayores de 60 años en Colombia

Algo que se ha venido repitiendo, por distintos medios de comunicación locales, es que el Coronavirus en 2021, debido a las mutaciones del virus, está afectando más gravemente y en mayor medida a personas más jóvenes.

Para ello se leen los conjuntos de datos, de los fallecidos por Covid-19 antes del primero de enero de 2021 y los muertos por dicha enfermedad en 2020.

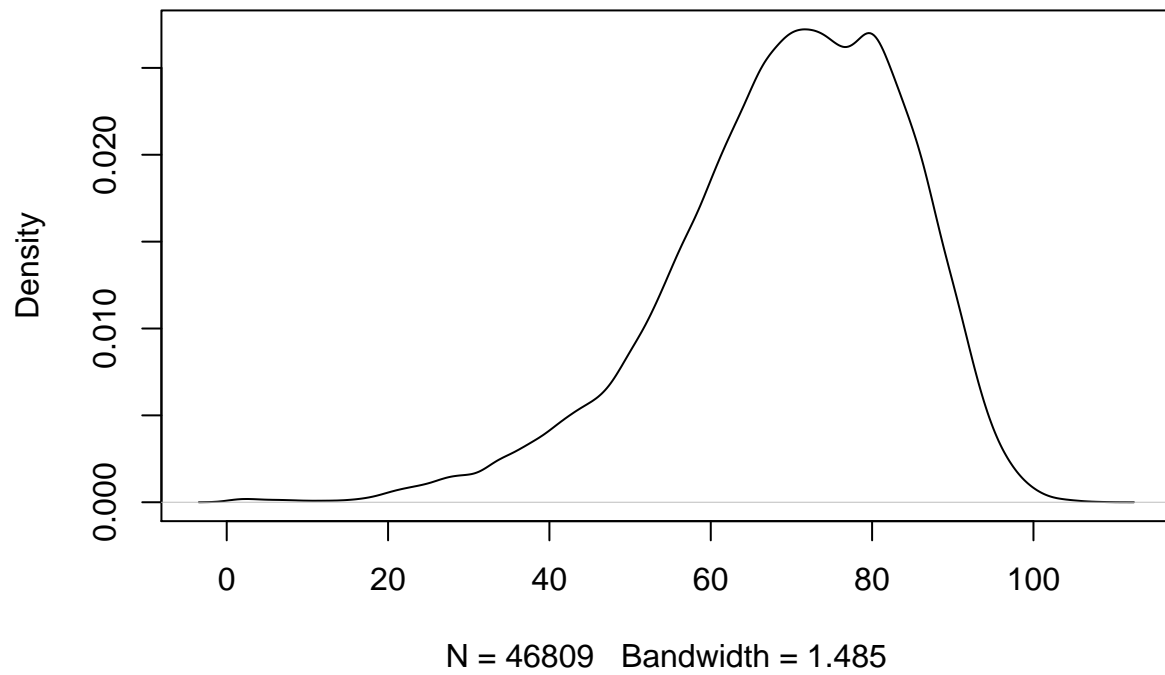
```
fallecidos_2020 <- read.csv("./out/fallecidos2020.csv")
fallecidos_2021 <- read.csv("./out/fallecidos2021.csv")
```

Análisis visual Para hacer un primer análisis visual, se hace plot de la distribución por edades de los fallecidos

```
edad_2020 <- as.vector(fallecidos_2020[, "Edad"])
edad_2021 <- as.vector(fallecidos_2021[, "Edad"])

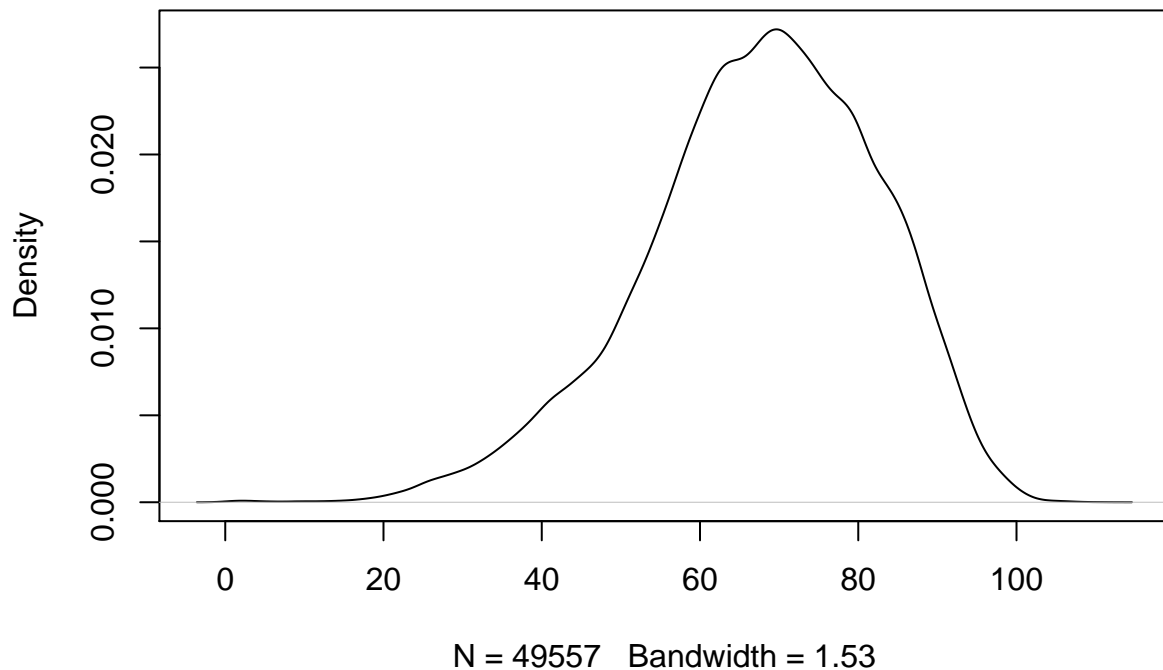
plot(density(edad_2020), main="Densidad de la edad en los fallecidos por Covid-19 - 2020")
```

Densidad de la edad en los fallecidos por Covid-19 – 2020



```
plot(density(edad_2021), main="Densidad de la edad en los fallecidos por Covid-19 - 2021")
```

Densidad de la edad en los fallecidos por Covid-19 – 2021



Ahora bien, realizamos las pruebas de normalidad, recordemos que estas pruebas toman como la hipótesis nula que las muestras provienen de una distribución normal con las distintas pruebas de normalidad que tiene R

```
ks.test(x=edad_2020,y='pnorm', mean=mean(edad_2020), sd=sd(edad_2020))
```

Kolmogorov-Smirnov

```
## Warning in ks.test(x = edad_2020, y = "pnorm", mean = mean(edad_2020), sd =  
## sd(edad_2020)): ties should not be present for the Kolmogorov-Smirnov test
```

```
##  
## One-sample Kolmogorov-Smirnov test  
##  
## data: edad_2020  
## D = 0.05812, p-value < 2.2e-16  
## alternative hypothesis: two-sided
```

```
ks.test(x=edad_2021,y='pnorm', mean=mean(edad_2021), sd=sd(edad_2021))
```

```
## Warning in ks.test(x = edad_2021, y = "pnorm", mean = mean(edad_2021), sd =  
## sd(edad_2021)): ties should not be present for the Kolmogorov-Smirnov test
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: edad_2021
## D = 0.041033, p-value < 2.2e-16
## alternative hypothesis: two-sided
```

Como podemos ver, por los resultados de las correspondientes pruebas de normalidad Kolmogorov-Smirnov, el valor de p es menor que 0.05, que por consenso es nuestro nivel de significancia para rechazar o aceptar la hipótesis nula. Si bien en las gráficas no eran lo suficientemente claras, la prueba K-S nos demuestra que no provienen, ninguna de las dos muestras, de una distribución normal.

Esta información es de vital importancia para escoger el tipo de técnica a usar en nuestra siguiente prueba.

Definición de la hipótesis Para afirmar que las personas más afectadas siguen siendo los adultos mayores de 60 años, deberíamos probar que la media de edad de los fallecidos en 2020 es mayor que 60 y que, a su vez, la media de edad de fallecidos en 2021 también.

Probando que la media de edad en los fallecidos en 2020 es mayor que 60 Como pudimos ver, no tan claramente en las gráficas pero sí con la prueba K-S, los datos de las edades de los fallecidos por Covid-19 en ambos años no son distribuciones normales, es por esta razón que utilizamos una prueba de hipótesis no paramétrica: Mann-Whitney-Wilcoxon que en R está predefinida como “wilcox.test”, recordando siempre que el criterio de rechazo de la hipótesis nula es que p-value sea menor que 0.05.

```
wilcox.test(edad_2020, mu=60, alternative = "g", paired = F, conf.level = 0.95)
```

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: edad_2020
## V = 869431729, p-value < 2.2e-16
## alternative hypothesis: true location is greater than 60
```

La hipótesis nula de la anterior prueba rezaba lo siguiente: La mediana de la población de la muestra es menor o igual que 60, y la hipótesis alternativa que la mediana de la población es mayor a 60, dado que p-value es menor que 0.05 se rechaza la hipótesis nula y se afirma la alternativa, es decir: la mayoría de los fallecidos son mayores de 60 años en el año 2020.

```
wilcox.test(edad_2021, mu=60, alternative = "g", paired = F, conf.level = 0.95)
```

Probando que la media de edad en los fallecidos en 2021 es mayor que 60

```
##
## Wilcoxon signed rank test with continuity correction
##
## data: edad_2021
## V = 908264529, p-value < 2.2e-16
## alternative hypothesis: true location is greater than 60
```

Al igual que en los datos recolectados de los fallecidos por Covid-19 en 2020, los datos de fallecidos por Covid-19 en 2021, nos indican, después de realizar la prueba de hipótesis, que la mayoría de las personas fallecidas son mayores de 60 años.

Conclusion de la afirmación en el infograma: Así pues, al tener comprobado que en ambos años la población más afectada por Covid-19, puesto que fallecieron, fueron los mayores de 60 años, podemos concluir que la población vulnerable sigue siendo esta y no la de los adultos jóvenes (menores de 60), como se dice en algunos medios de comunicación y en rumores de redes sociales.

Afirmación 2: Los contagios y muertes en 2021 se han mantenido o venido en aumento, incluso después del inicio de la vacunación (marzo 2021), en Colombia

Aquí hacemos uso de la base de datos recolectado de manera propia, tomando los tweets de la cuenta oficial del Ministerio de la Salud en Twitter que reportaban los nuevos contagios y fallecidos día a día en Colombia, también pre-procesados en Python3

Empezamos con la lectura de los datos

```
reportes_antes_vacunacion <-read.csv("./out/reportes_antes_vacunacion_minsalud.csv")
reportes_despues_vacunacion <-read.csv("./out/reportes_despues_vacunacion_minsalud.csv")
```

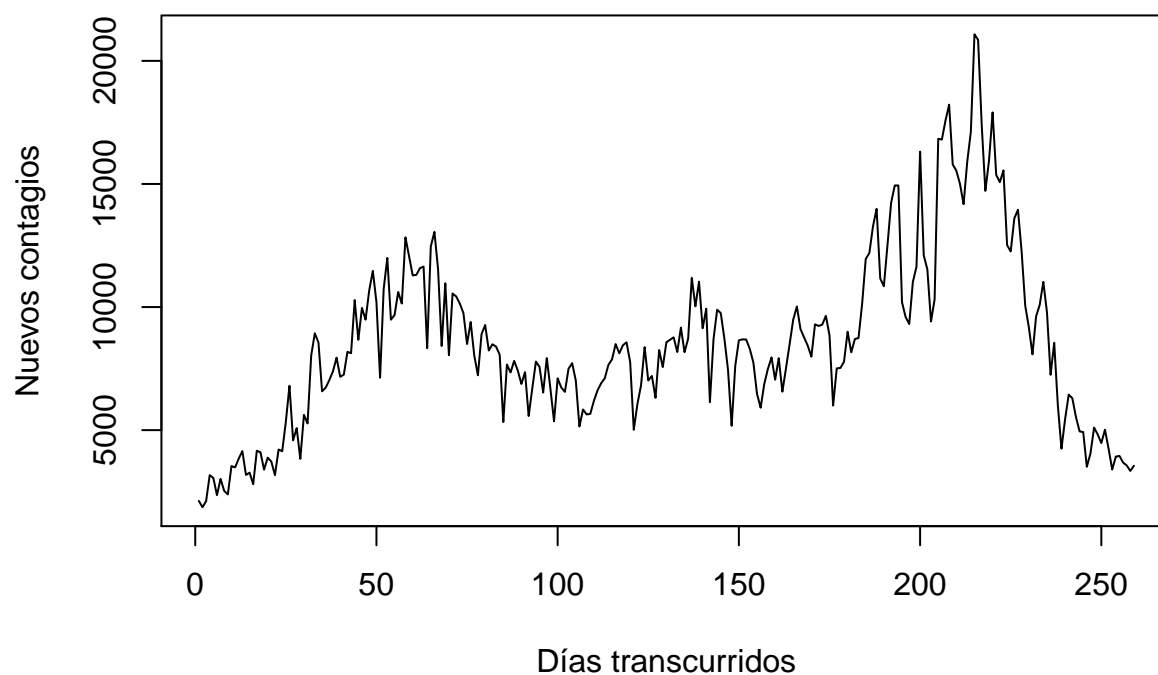
Una breve explicación de las columnas:

- recuperados: Número de pacientes diagnosticados con Covid-19 recuperados y registrados ese día
- nuevos: Número de nuevas personas diagnosticados con Covid-19 y registrados ese día
- fallecidos: Número de personas fallecidas por Covid-19 registrados ese día
- muestras: Número de muestras para testear Covid-19 tomadas ese día
- date: Fecha de la toma del reporte
- Cada fila de los datasets corresponde al resumen de las columnas en dicho día.
- El primer día para los reportes_2020, es decir la primer fila, corresponde al 15 de junio de 2020 y la última fila al 31 de diciembre de 2020
- El primer día para los reportes_2021, es decir la primer fila, corresponde al 1 de enero de 2021 y la última fila al 13 de junio de 2021

Análisis visual previo de los contagios:

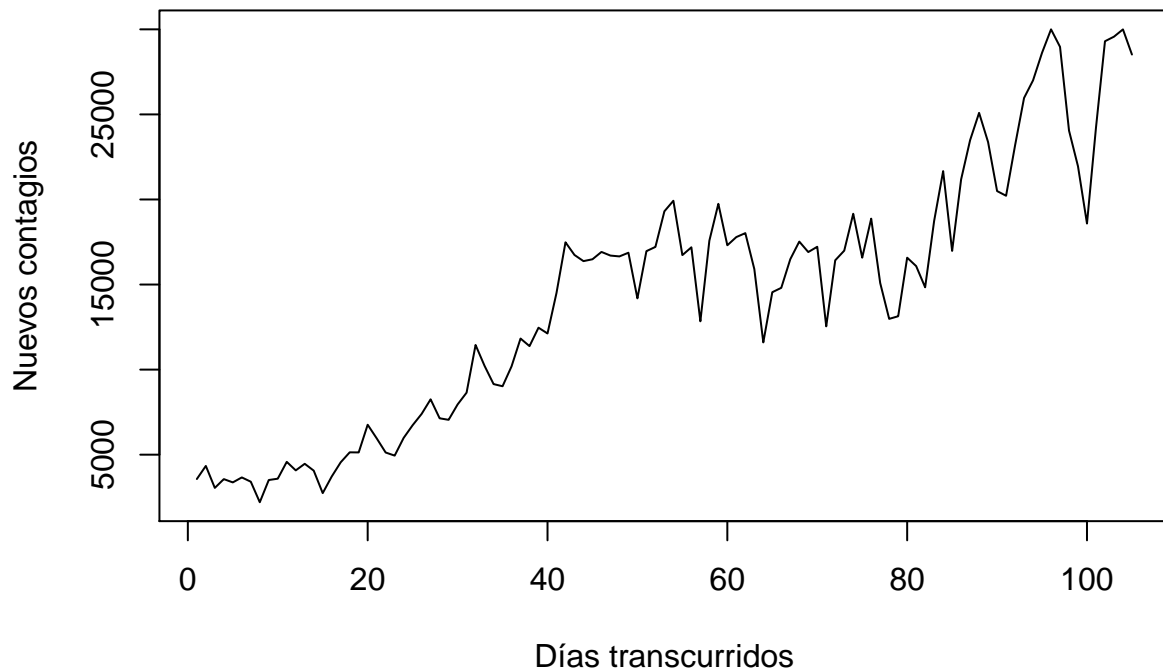
```
plot(reportes_antes_vacunacion$nuevos, type="l", main="Reportes diarios de nuevos contagios Covid-19")
```

Reportes diarios de nuevos contagios Covid-19 antes de la vacunaci



```
plot(reportes_despues_vacunacion$nuevos, type="l", main="Reportes diarios de nuevos contagios Covid-19")
```

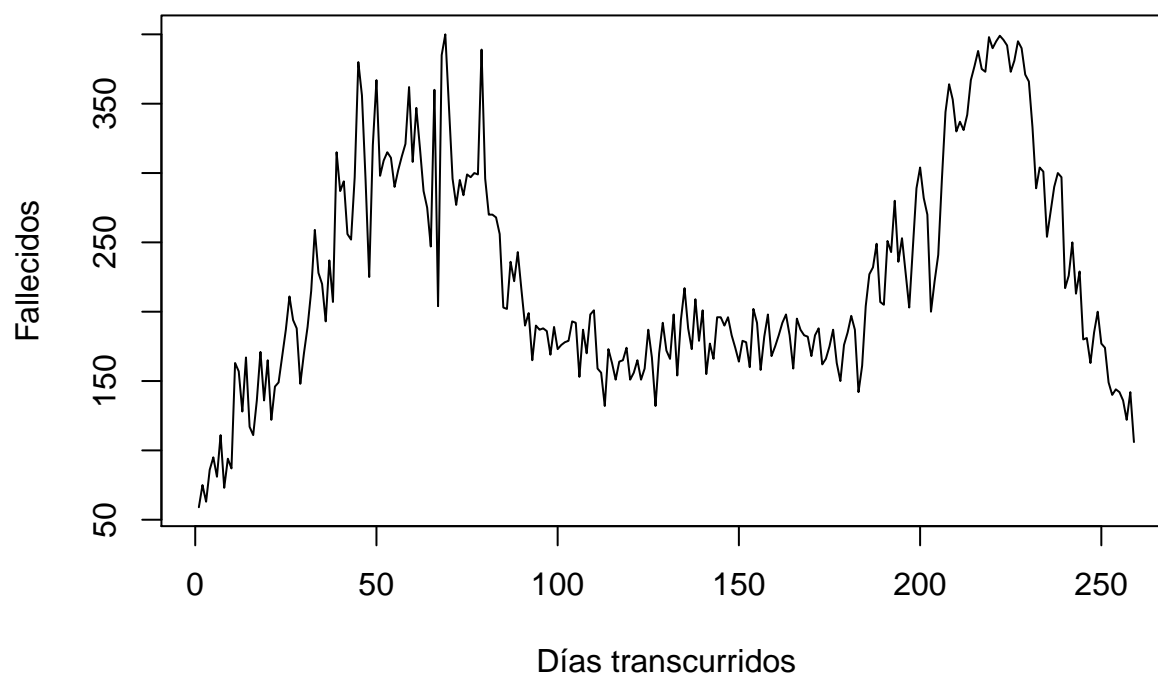
Reportes diarios de nuevos contagios Covid-19 después de la vacuna



Análisis visual previo de los fallecimientos:

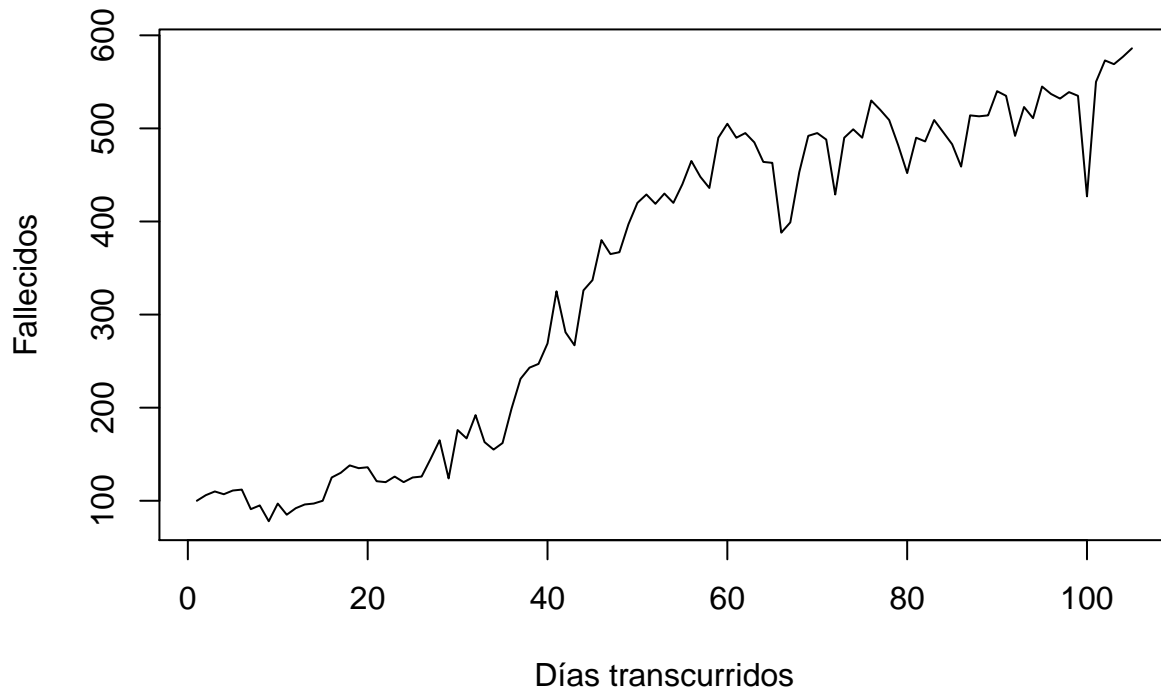
```
plot(reportes_antes_vacunacion$fallecidos, type="l", main="Reportes diarios de fallecidos por Covid-19")
```

Reportes diarios de fallecidos por Covid-19 antes de la vacunación



```
plot(reportes_despues_vacunacion$fallecidos, type="l", main="Reportes diarios de fallecidos por Covid-19")
```


Reportes diarios de fallecidos por Covid-19 despues de la vacunaci



Como se desea demostrar algo sobre los contagios y muertes de los diferentes momentos, se debe dividir dichas hipotesis en 2, una para contagios y otra para muertes.

Definicion de hipotesis 1: ¿Han disminuido los contagios desde el inicio de la vacunacion? Podemos traducirlo en probar si en los días anteriores a la vacunación hubo menos contagiados que en los días posteriores, de esto ser cierto, querría decir que en los días posteriores o se mantuvo el ritmo de contagios o aumentó. (Negando la afirmación de que disminuyeron)

¿Que distribución tienen estos datos? Primero se debe saber que tipo de distribución tienen los datos (contagios por día antes y despues de la vacunación), para escoger una estadística paramétrica adecuada, por lo tanto realizamos la prueba de normalidad Kolmogorov-Smirnov

```
contagios_antes <-reportes_antes_vacunacion$nuevos
contagios_despues <-reportes_despues_vacunacion$nuevos
ks.test(x=contagios_antes,y='pnorm', mean=mean(contagios_antes), sd=sd(contagios_antes))
```

```
## Warning in ks.test(x = contagios_antes, y = "pnorm", mean =
## mean(contagios_antes), : ties should not be present for the Kolmogorov-Smirnov
## test
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: contagios_antes
```

```
## D = 0.091429, p-value = 0.02633
## alternative hypothesis: two-sided
```

```
ks.test(x=contagios_despues,y='pnorm', mean=mean(contagios_antes), sd=sd(contagios_despues))
```

```
##
## One-sample Kolmogorov-Smirnov test
##
## data: contagios_despues
## D = 0.35103, p-value = 1.157e-11
## alternative hypothesis: two-sided
```

Siendo la hipótesis nula de Kolmogorov-Smirnov que los datos provienen de una distribución normal y $p\text{-value} < 0.05$, ésta debe ser rechazada, por lo tanto los datos no tienen distribución normal y no pueden ser tratados con técnicas de prueba de hipótesis paramétricas.

Por lo anterior se utiliza Wilcoxon (U Mann Whitney) para probar las hipótesis.

Prueba de la hipótesis:

- H_0 : Los contagios aumentaron o se mantuvieron iguales en ambas muestra | $Me_1 \geq Me_2$
- H_1 : Hubo menos contagios antes de la vacunación | $Me_1 < Me_2$

```
nA <- length(contagios_antes)
nD <- length(contagios_despues)
muestras <- c(rep(1, nA), rep(2, nD))
contagios <- c(contagios_antes, contagios_despues)
#Prueba de U Mann-Whitney
wilcox.test(contagios~muestras, paired=F, mu=0, alternative="l", conf.level = "0.95")
```

```
##
## Wilcoxon rank sum test with continuity correction
##
## data: contagios by muestras
## W = 7529.5, p-value = 1.268e-11
## alternative hypothesis: true location shift is less than 0
```

El p-value devuelto por esta prueba no-paramétrica nos indican que debemos rechazar la hipótesis nula, es decir, podemos concluir que: la mediana de contagios antes del inicio de la vacunación es menor a la mediana de contagiados después de el inicio de esta. Por lo tanto, la afirmación de que los contagios aumentaron, incluso empezado el plan de vacunación en marzo es verdadera.

Definición de hipótesis 2: ¿Han disminuido las muertes desde el inicio de la vacunación? Podemos traducirlo en probar si en los días anteriores a la vacunación hubo menos muertos que en los días posteriores, de esto ser cierto, querría decir que en los días posteriores o se mantuvo el ritmo de contagios o aumentó. (Negando la afirmación de que disminuyeron)

¿Que distribución tienen estos datos? Primero se debe saber que tipo de distribución tienen los datos (fallecidos por día antes y después de la vacunación), para escoger una estadística paramétrica adecuada, por lo tanto realizamos la prueba de normalidad Kolmogorov-Smirnov

```

muertos_antes <-reportes_antes_vacunacion$fallecidos
muertos_despues <-reportes_despues_vacunacion$fallecidos
ks.test(x=muertos_antes,y='pnorm', mean=mean(muertos_antes), sd=sd(muertos_antes))

## Warning in ks.test(x = muertos_antes, y = "pnorm", mean = mean(muertos_antes), :
## ties should not be present for the Kolmogorov-Smirnov test

##
## One-sample Kolmogorov-Smirnov test
##
## data: muertos_antes
## D = 0.15701, p-value = 5.691e-06
## alternative hypothesis: two-sided

ks.test(x=muertos_despues,y='pnorm', mean=mean(contagios_antes), sd=sd(muertos_despues))

## Warning in ks.test(x = muertos_despues, y = "pnorm", mean =
## mean(contagios_antes), : ties should not be present for the Kolmogorov-Smirnov
## test

##
## One-sample Kolmogorov-Smirnov test
##
## data: muertos_despues
## D = 1, p-value < 2.2e-16
## alternative hypothesis: two-sided

```

Siendo la hipótesis nula de Kolmogorov-Smirnov que los datos provienen de una distribución normal y $p\text{-value} < 0.05$, está debe ser rechazada, por lo tanto los datos no tienen distribución normal y no pueden ser tratados con técnicas de prueba de hipótesis paramétricas.

Por lo anterior se utiliza Wilcoxon (U Mann Whitney) para probar las hipótesis.

Prueba de la hipótesis:

- H_0 : Los muertos aumentaron o se mantuvieron iguales en ambas muestra | $Me_1 \geq Me_2$
- H_1 : Hubo menos contagios antes de la vacunación | $Me_1 < Me_2$

```

nA <- length(muertos_antes)
nD <- length(muertos_despues)
muestras <- c(rep(1, nA), rep(2, nD))
muertos <- c(muertos_antes, muertos_despues)
#Prueba de U Mann-Whitney
wilcox.test(muertos~muestras, paired=F, mu=0, alternative="l", conf.level = "0.95")

##
## Wilcoxon rank sum test with continuity correction
##
## data: muertos by muestras
## W = 9071.5, p-value = 3.247e-07
## alternative hypothesis: true location shift is less than 0

```

El p-value devuelto por esta prueba no-parametrica nos indican que debemos rechazar la hipotesis nula, es decir, podemos concluir que: la mediana de muertos antes del inicio de la vacunación es menor a la mediana de muertos despues de el inicio de esta. Por lo tanto, la afirmación de que los muertos por Covid-19 aumentaron, incluso empezado el plan de vacunación en marzo es verdadera.

Conclusión de la afirmación en el infograma: Se probó, mediante pruebas estadísticas que los niveles de contagio (hipotesis 1) y de fallecimientos por Covid-19 (hipotesis 2) eran menores en los días anteriores al inicio de la vacunación (marzo de 2021), es decir se han incrementado los efectos del Covid-19 en Colombia incluso habiendo empezado el plan de vacunación.

Afirmación 3: Los contagios no aumentaron a causa del Paro Nacional, iniciado el 28 de abril de 2021

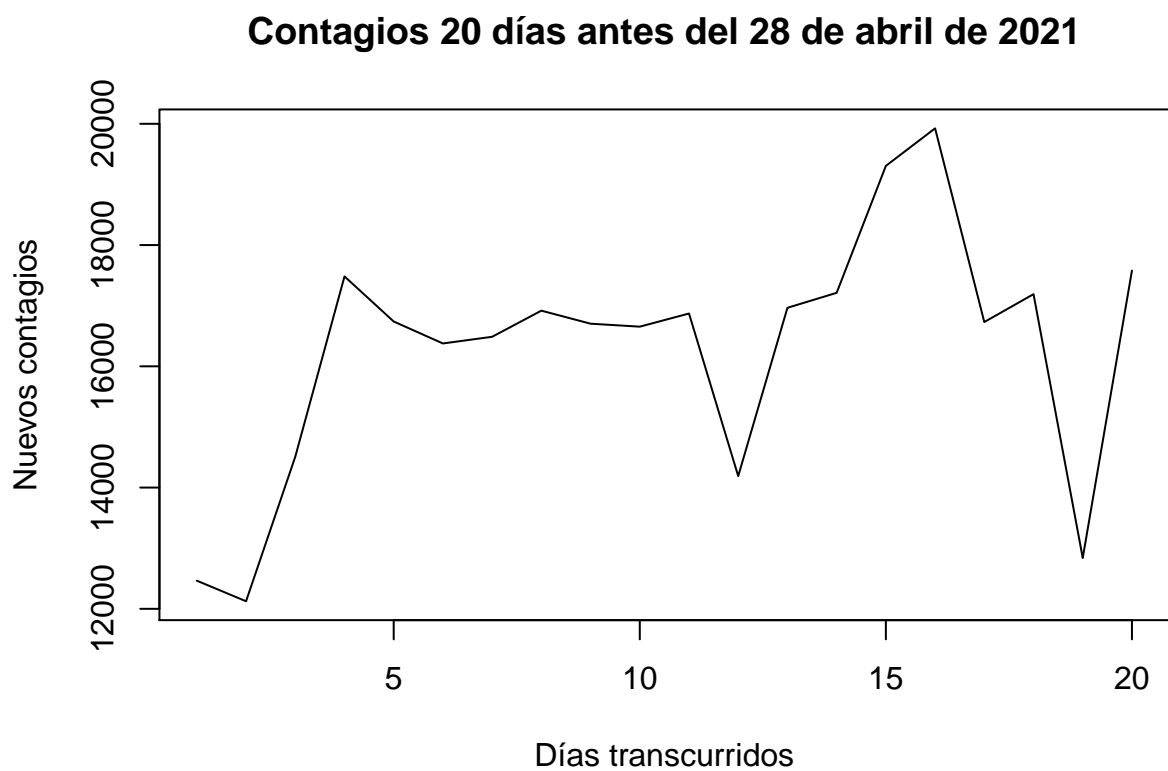
Según los expertos, el Corona Virus tiene un desarrollo dentro del cuerpo humano que dura, como máximo, 15 días, también que las aglomeraciones potencian la aceleración de contagio de este virus, sin embargo, en espacios abiertos y con medidas de bioseguridad baja la probabilidad de contagio. Algunos sectores políticos querían impedir las masivas movilizaciones aludiendo que dichas aglomeraciones complicarían la situación de contagios por Covid-19 en Colombia.

Se empieza con la lectura de las bases de datos, también pre-procesadas con Python. Se van a leer dos muestras, de los reportes diarios que hace el Ministerio de la Salud en su cuenta de twitter, de las cuales la primer muestra corresponde a los reportes de los 20 días antes del 28 de abril y la segunda muestra a los 20 días siguientes al 28 de abril, día en que dió inicio el Paro Nacional y las masivas manifestaciones.

```
antes_28a <- read.csv("./out/reportes_antes_28a_minsalud.csv")
despues_28a <-read.csv("./out/reportes_despues_28a_minsalud.csv")
contagios_antes_28a <- antes_28a$nuevos
contagios_despues_28a <- despues_28a$nuevos
```

Análisis visual de los datos

```
plot(contagios_antes_28a, type="l", main="Contagios 20 días antes del 28 de abril de 2021", xlab="Días "
```



```
plot(contagios_despues_28a, type="l", main="Contagios 20 días después del 28 de abril de 2021", xlab=
```



¿Qué distribución tiene? En el análisis visual, podríamos concluir que ninguna de las dos muestras proviene de una distribución normal, sin embargo, por protocolo realizaremos la prueba de normalidad.

Son 20 filas en cada muestra, lo que nos permite hacer uso de la prueba de normalidad Shapiro

```
shapiro.test(contagios_antes_28a)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  contagios_antes_28a  
## W = 0.88341, p-value = 0.02039
```

```
shapiro.test(contagios_despues_28a)
```

```
##  
## Shapiro-Wilk normality test  
##  
## data:  contagios_despues_28a  
## W = 0.9509, p-value = 0.381
```

Podemos observar que el p-value para la muestra de los reportes antes del 28 de abril es menor que 0.05, por esto rechazamos la hipótesis nula, que en el caso de shapiro, corresponde a que la muestra proviene de una distribución normal, es decir no podemos asumir la normalidad de dicha muestra.

Por otra parte, tenemos que el p-value, para los reportes después del 28 de abril es mayor que 0,05, por lo que debemos rechazar la hipótesis nula y asumir que dichos datos provienen de una distribución normal.

Al tener distribuciones mixtas podemos hacer uso, de nuevo, de la prueba Wilcoxon, en este caso de paridad y de rangos con signo, porque las muestras son dependientes, dado que están conectadas a un suceso entre ellas dos, como lo es en la mayoría de ejemplos de la literatura la aplicación de un fármaco o tratamiento.

Definición de hipótesis: ¿Impactaron las marchas y aglomeraciones a nivel nacional, que empezaron el 28 de abril en Colombia, en la cifra de contagios por Covid?

Tomando en cuenta que las muestras son de 20 días antes y 20 días después al inicio de las marchas, si estas hubiesen aumentado el nivel de contagios entonces las medias poblacionales deberían ser distintas y las posteriores al 28 de abril mayores a las anteriores. Basta con hacer una prueba de Mann-Whitney-Wilcoxon para mirar si las muestras son significativamente diferentes.

Prueba de hipótesis:

- H0: Las medias poblacionales son significativamente similares
- H1: Las medias poblacionales no son significativamente similares

además se debe usar el parámetro de **paired**, porque las muestras son dependientes una de la otra

```
wilcox.test(x=contagios_antes_28a, y=contagios_despues_28a, alternative = "t", mu=0, paired=T)
```

```
##  
## Wilcoxon signed rank test  
##  
## data: contagios_antes_28a and contagios_despues_28a  
## V = 108, p-value = 0.9273  
## alternative hypothesis: true location shift is not equal to 0
```

Aplicando la prueba de Wilcoxon con las muestras, obtenemos un p-value mayor a 0.05, por lo que no podemos rechazar la hipótesis nula, es decir, las muestras son significativamente similares, por lo tanto concluimos que los eventos del 28 de abril y días posteriores, como aglomeraciones y eventos sociales masivos no influyeron significativamente en el aumento de contagios por Covid-19 en Colombia.

Conclusiones de afirmación; Pudimos demostrar que **NO** habían diferencias significativas dentro de las muestras, de contagios 20 días antes y 20 días después del 28 de abril, por lo tanto podemos concluir que no hay un incremento causado por los eventos ocurridos entre el 28 de abril y -aproximadamente- 15 días después de esto, pues tomando como axioma que la enfermedad afecta alrededor de 15 días y que las personas presentan síntomas o son diagnosticadas 5 días después de su contagio, de haberse contagiado en los eventos de esos días, los registros hubiesen aumentado más de lo que lo hicieron en ese margen de tiempo.

Afirmación 4: Las muertes por Covid-19 si aumentaron después del 28 de abril de 2021:

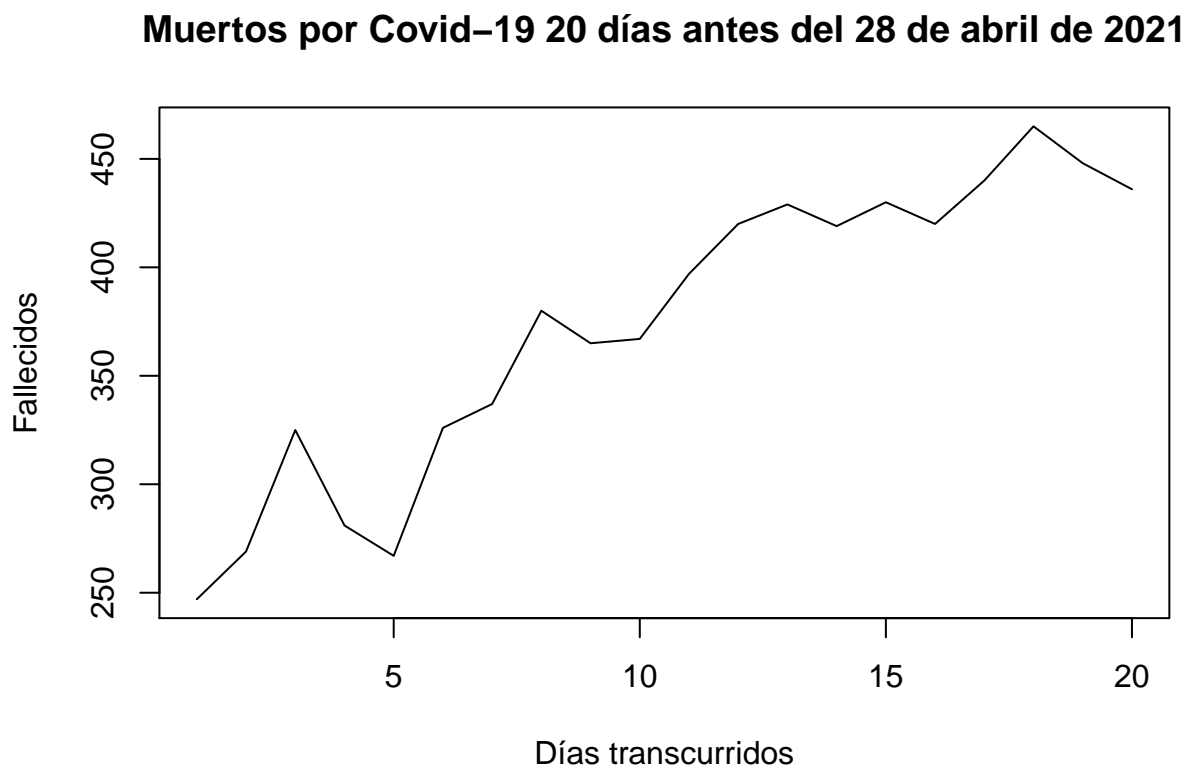
Si bien se demostró en la afirmación anterior que los contagios no aumentaron después del 28 de abril en Colombia, los problemas de orden público si pudieron haber influido en la cantidad de muertos por Covid-19, puesto que los heridos de las protestas sociales o la ausencia de suministros para el tratamiento del Covid-19 si pudieron haber influenciado en la ocupación o tratamientos de los pacientes de UCI y por ende la muerte en algunos de ellos.

De nuevo se empieza con la lectura de los conjuntos de datos, en la afirmación anterior ya habíamos tomado los reportes del Ministerio de la Salud 20 días antes y 20 días después del 28 de abril, por tal razón en esta sección no vamos a leer un nuevo .csv, sino que vamos a extraer la columna “fallecidos” de lo que ya se cargó anteriormente

```
muertos_antes_28a <- antes_28a$fallecidos  
muertos_despues_28a <- despues_28a$fallecidos
```

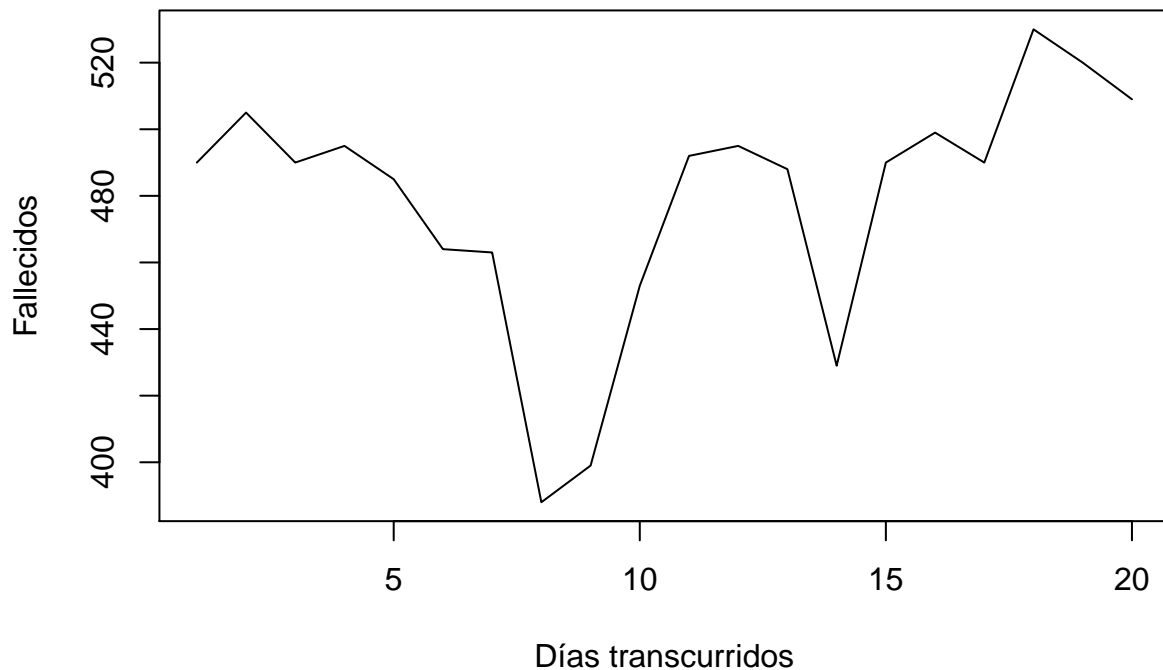
Análisis visual de los datos

```
plot(muertos_antes_28a, type="l", main = "Muertos por Covid-19 20 días antes del 28 de abril de 2021", xlab = "Días transcurridos", ylab = "Fallecidos")
```



```
plot(muertos_despues_28a, type="l", main = "Muertos por Covid-19 20 días después del 28 de abril de 2021", xlab = "Días transcurridos", ylab = "Fallecidos")
```


Muertos por Covid-19 20 días después del 28 de abril de 2021



Visualmente es notorio que las muestras no siguen una distribución normal, sin embargo, debemos hacer las correspondientes pruebas de normalidad, para así tomar la prueba estadística que mejor vaya con las características de los datos, de nuevo, al ser solo 20 filas para cada muestra, podemos hacer uso de Shapiro Test en vez de la Kolmogorov-Smirnov

```
shapiro.test(muertos_antes_28a)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  muertos_antes_28a  
## W = 0.91048, p-value = 0.06509
```

```
shapiro.test(muertos_despues_28a)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  muertos_despues_28a  
## W = 0.85735, p-value = 0.007095
```

Siguiendo nuestro consenso de $\alpha=0.05$ tenemos que el p-value para la muestra de fallecidos antes del 28 de abril es de 0.07, lo que indica que es mayor a α y en nuestro criterio de rechazo no podemos invalidar la hipótesis nula de Shapiro, lo que se traduce en que no podemos descartar que la muestra proviene de una distribución NORMAL, por lo tanto, asumimos que es normal.

Por otra parte para los fallecidos despues del 28 de abril, el p-value es menor que 0.05 (0.007) lo que, según nuestro criterio de rechazo, nos obliga a rechazar la hipotesis nula y asumir la hipotesis alternativa, en este caso: la distribución normal no describe a esa muestra de datos.

Definicion de hipotesis: ¿Impactaron los eventos del 28 de abril en las muertes por Covid-19?

Siendo esta una incognita bastante ambigua, vamos a asumir que dichos eventos estuvieron relacionados con las muertes de pacientes de Covid-19 por la posible ocupación, imprevista, de camas UCI con pacientes heridos en las protestas y la falta de oxigeno y otros medicamentos, debido a bloqueos y desabastecimiento de los hospitales a causa del Paro Nacional, de nuevo teniendo en cuenta que el Covid-19 mata a sus enfermos graves en 15 días o menos.

Así pues, se procede a reformar la pregunta: ¿Incrementaron significativamente los muertos despues del 28 de abril de 2021 por causa de Covid-19?

Prueba de hipotesis:

- H0: Los fallecidos, por Covid-19, en los 20 días posteriores al 28 de abril fueron mayores, o iguales, que los 20 días anteriores al 28 de abril
- H1: Los fallecidos, por Covid-19, en los 20 días posteriores al 28 de abril fueron menores que los 20 días anteriores al 28 de abril

```
wilcox.test(x=muertos_despues_28a, y=muertos_antes_28a, alternative = "l", mu=0, paired=T)
```

```
##  
## Wilcoxon signed rank test  
##  
## data: muertos_despues_28a and muertos_antes_28a  
## V = 210, p-value = 1  
## alternative hypothesis: true location shift is less than 0
```

Como observamos, al ejecutar la prueba Wilcoxon, para muestras dependientes y con las hipotesis planteadas anteriormente, obtenemos p-value > 0.05 , según nuestro criterio de rechazo no debemos rechazar la hipotesis nula y podemos asumirla, en este caso podemos asumir que “Los fallecidos, por Covid-19, en los 20 días posteriores al 28 de abril fueron mayores, o iguales, que los 20 días anteriores al 28 de abril”. Sin embargo, nuestra afirmación consiste en que los muertos aumentaron, no que se mantuvieron, es por esta razón que debemos hacer la prueba de igualdad en las muestras

```
wilcox.test(x=muertos_despues_28a, y=muertos_antes_28a, alternative = "t", mu=0, paired=T)
```

```
##  
## Wilcoxon signed rank test  
##  
## data: muertos_despues_28a and muertos_antes_28a  
## V = 210, p-value = 1.907e-06  
## alternative hypothesis: true location shift is not equal to 0
```

Aquí obtenemos un p-value < 0.05 , por lo que debemos rechazar la hipotesis nula (que las muestras eran significativamente similares) y asumir la hipotesis alternativa: que las muestras son significativamente diferentes.

Conclusión de la afirmación: En las pruebas de la hipótesis planteadas, se demostró, en primera instancia, que los fallecidos por Covid-19 20 días después del 28 de abril de 2021 eran mayores o iguales que los fallecidos por Covid-19 20 días antes del 28 de abril, y se demostró, en una segunda prueba, que los fallecidos por Covid-19 20 días después del 28 de abril eran significativamente diferentes que los fallecidos por Covid-19 20 días antes del 28 de abril de 2021.

Por lo anteriormente dicho, podemos concluir que los fallecidos, por Covid-19, en el lapso del 29 de abril al 18 de mayo fueron significativamente mayores que los fallecidos, por Covid-19, en los 20 días anteriores al 28 de abril, en resumidas palabras, las muertes por Covid-19 aumentaron desde el 28 de abril de 2021 con respecto a los 20 días anteriores.