Primer parcial

Estudiantes

Cristian Alberto Cortes Zarate
Santiago Carvajal Torres

Docente

Mauricio Alejandro Mazo Lopera

Asignatura

Series de Tiempo Univariadas

UNIVERSIDAD
NACIONAL
DE COLOMBIA

Sede Medellín
13 de septiembre de 2022

Índice

1. Primer punto.

$$X_t = W_{t-2} + 0.5W_{t-1} + 2W_t + 0.5W_{t+1} + W_{t+2}$$

Donde los W_t son independientes con $\mathbf{E}[W_t] = 0$ y varianza $\sigma_w^2 = 4.8$

a. Encuentre la media y la varianza del proceso.

Media

$$\begin{aligned}
&\mathbf{E}[X_t] = \mathbf{E}[W_{t-2} + 0.5W_{t-1} + 2W_t + 0.5W_{t+1} + W_{t+2}] \\
&\mathbf{E}[X_t] = \mathbf{E}[W_{t-2}] + 0.5\mathbf{E}[W_{t-1}] + 2\mathbf{E}[W_t] + 0.5\mathbf{E}[W_{t+1}] + \mathbf{E}[W_{t+2}] \\
&\mathbf{E}[X_t] = 0 + 0.5(0) + 2(0) + 0.5(0) + 0 \\
&\mathbf{E}[X_t] = 0
\end{aligned}$$

Varianza

$$\begin{aligned}
&\mathbf{V}ar[X_t] = \mathbf{V}ar[W_{t-2} + 0.5W_{t-1} + 2W_t + 0.5W_{t+1} + W_{t+2}] \\
&\mathbf{V}ar[X_t] = \mathbf{V}ar[W_{t-2}] + \mathbf{V}ar[0.5W_{t-1}] + \mathbf{V}ar[2W_t] + \mathbf{V}ar[0.5W_{t+1}] + \mathbf{V}ar[W_{t+2}] \\
&\mathbf{V}ar[X_t] = \mathbf{V}ar[W_{t-2}] + 0.5^2\mathbf{V}ar[W_{t-1}] + 2^2\mathbf{V}ar[W_t] + 0.5^2\mathbf{V}ar[W_{t+1}] + \mathbf{V}ar[W_{t+2}] \\
&\mathbf{V}ar[X_t] = \sigma_w^2 + 0.25\sigma_w^2 + 4\sigma_w^2 + 0.25\sigma_w^2 + \sigma_w^2 \\
&\mathbf{V}ar[X_t] = 6.5\sigma_w^2 \\
&\mathbf{V}ar[X_t] = 6.5(4.8) \\
&\mathbf{V}ar[X_t] = 31.2
\end{aligned}$$

b.Encuentre y grafique las funciones ACF y PACF del proceso.

ACF

$$\begin{aligned}
&\gamma(t, t-h) = Cov[X_t, X_{t-h}] \\
&= Cov[W_{t-2} + 0.5W_{t-1} + 2W_t + 0.5W_{t+1} + W_{t+2}, W_{t-h-2} + 0.5W_{t-h-1} + 2W_{t-h} + 0.5W_{t-h+1} + W_{t-h+2}] \\
&= Cov[W_{t-2}, W_{t-h-2}] + 0.5Cov[W_{t-2}, W_{t-h-1}] + 2Cov[W_{t-2}, W_{t-h}] + 0.5Cov[W_{t-2}, W_{t-h+1}] \\
&+ Cov[W_{t-2}, W_{t-h+2}] + 0.5Cov[W_{t-1}, W_{t-h-2}] + 0.25Cov[W_{t-1}, W_{t-h-1} + Cov[W_{t-1}, W_{t-h}] \\
&+ 0.25Cov[W_{t-1}, W_{t-h+1}] + Cov[W_{t-1}, W_{t-h+2}] + 2Cov[W_t, W_{t-h-2}] + Cov[W_t, W_{t-h-1}] \\
&+ 4Cov[W_t, W_{t-h}] + Cov[W_t, W_{t-h+1}] + 2Cov[W_t, W_{t-h+2}] + 0.5Cov[W_{t+1}, W_{t-h-2}] \\
&+ 0.25Cov[W_{t+1}, W_{t-h-1}] + Cov[W_{t+1}, W_{t-h}] + 0.25Cov[W_{t+1}, W_{t-h+1}] + 0.5Cov[W_{t+1}, W_{t-h+2}] \\
&+ Cov[W_{t+2}, W_{t-h-2}] + 0.5Cov[W_{t+2}, W_{t-h-1}] + 2Cov[W_{t+2}, W_{t-h}] + 0.5Cov[W_{t+2}, W_{t-h+1}] \\
&+ Cov[W_{t+2}, W_{t-h+2}]]
\end{aligned}$$

Con h=1

$$\begin{aligned}\gamma(t,t-1) &= 0.5Cov[W_{t-2},W_{t-2}] + Cov[W_{t-1},W_{t-1}] + Cov[W_t,W_t] + 0.5Cov[W_{t+1},W_{t+1}] \\ &= 0.5\sigma^2 + \sigma_w^2 + \sigma_w^2 + 0.5\sigma_w^2 = 3\sigma_w^2\end{aligned}$$

Con h=2

$$\begin{aligned}\gamma(t,t-2) &= 2Cov[W_{t-2},W_{t-2}] + 0.25Cov[W_{t-1},W_{t-1}] + 2Cov[W_t,W_t] \\ &= 2\sigma^2 + 0.25\sigma_w^2 + 2\sigma_w^2 = 4.25\sigma_w^2\end{aligned}$$

Con h=3

$$\begin{aligned}\gamma(t,t-3) &= 0.5Cov[W_{t-2},W_{t-2}] + 0.25Cov[W_{t-1},W_{t-1}] \\ &= 0.5\sigma_w^2 + \sigma_w^2 = 1.5\sigma_w^2\end{aligned}$$

Con h=4

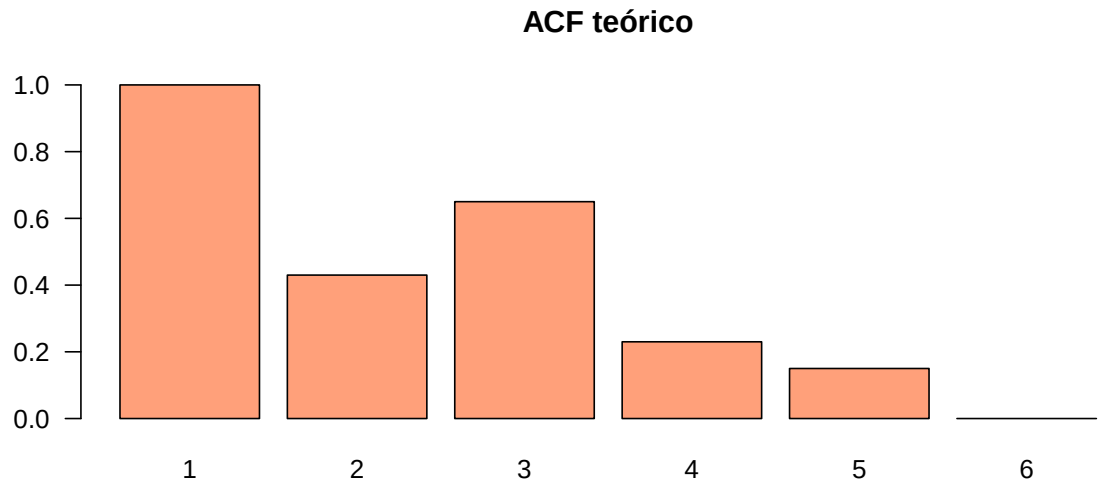
$$\begin{aligned}\gamma(t,t-4) &= Cov[W_{t-2},W_{t-2}] \\ &= \sigma_w^2\end{aligned}$$

Con h=5

$$\gamma(t,t-5) = 0$$

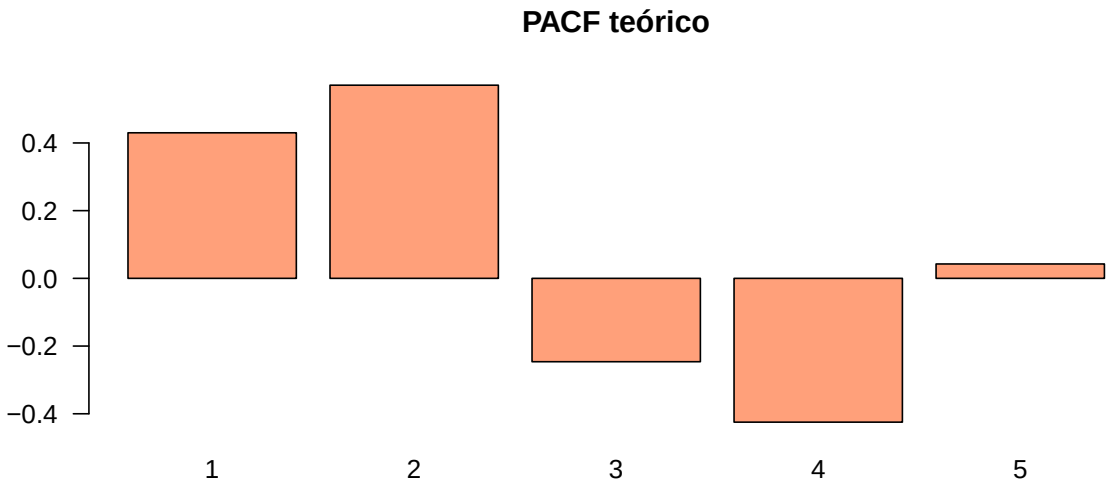
Con h ≥ 5 es 0

```
acf <- c(1,0.43,0.65,0.23,0.15,0)
```



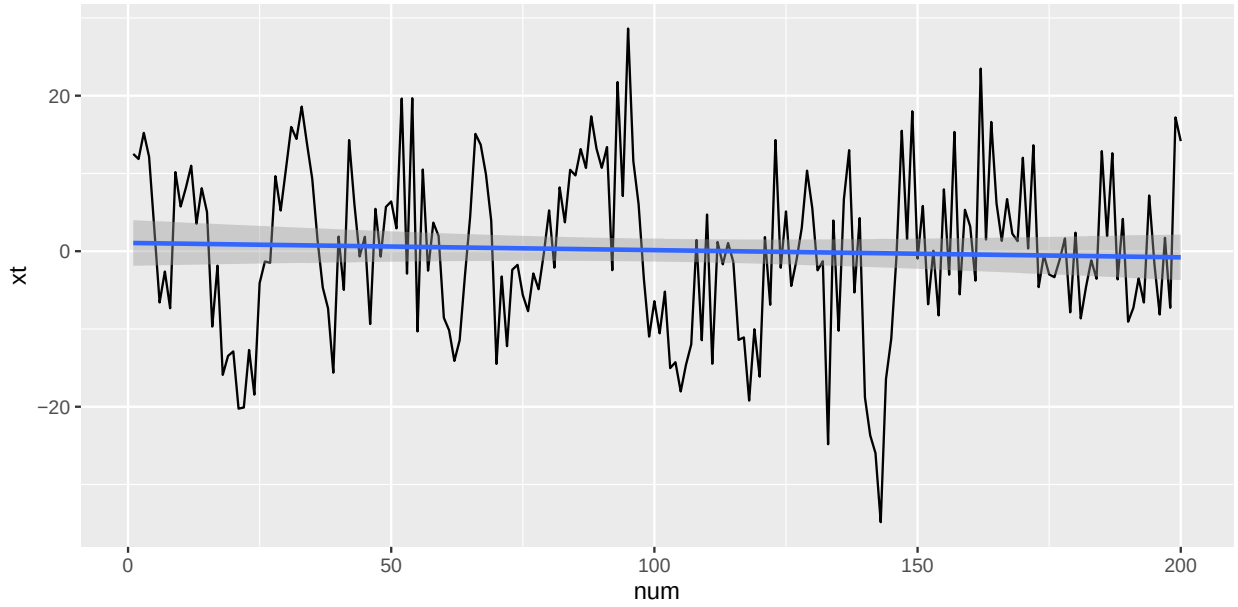
PACF

```
acf_ma1<-c(0.43,0.65,0.23,0.15,0)
pacf_ma1<-vector()
pacf_ma1[1]<-acf_ma1[1]
for (i in 2:5){
  deno<-toeplitz(c(1,acf_ma1[1:(i-1)]))
  aux_1<-deno
  aux_1[,i]<-acf_ma1[1:i]
  nume<-aux_1
  pacf_ma1[i]<-det(nume)/det(deno)}
```

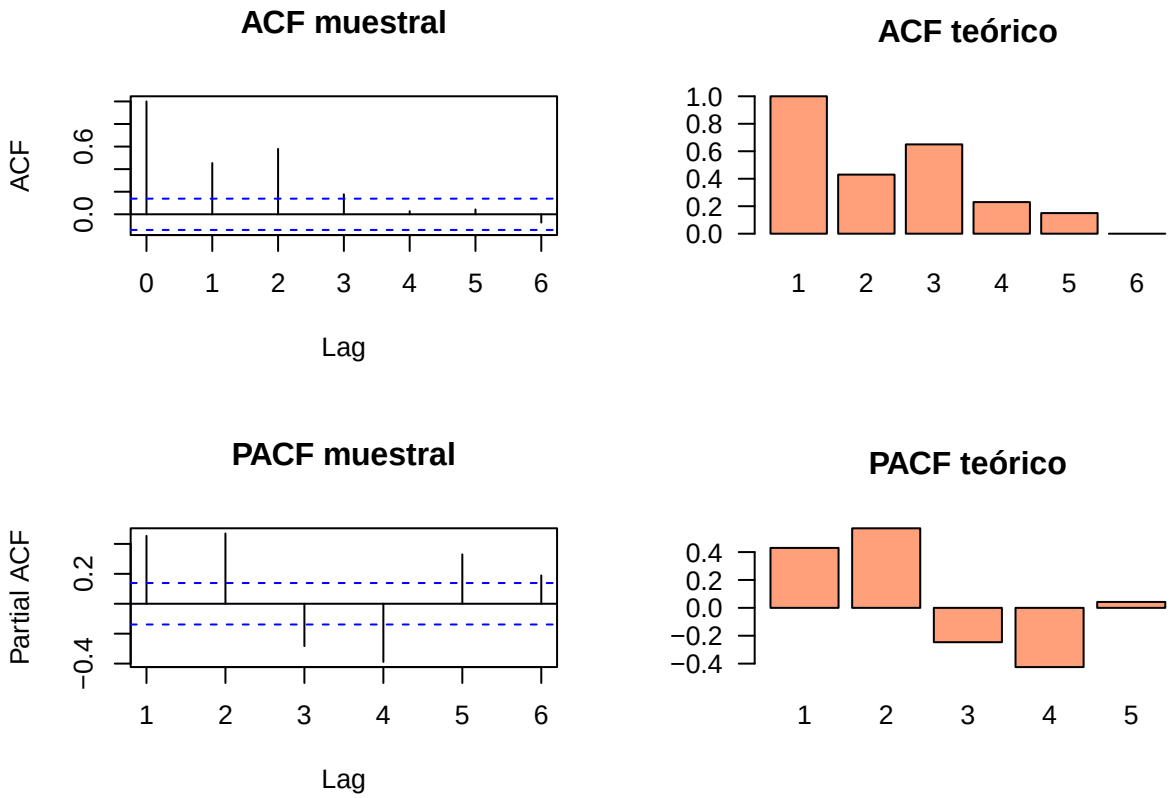


c. Simule y grafique una realización del proceso estocástico X t de tamaño 200, dando los valores que usted considere apropiados a cada uno de los parámetros.

```
sigma_w <- 4.8
wt <- rnorm(204,0,sigma_w)
xt <- vector()
for (t in 3:203) {
  xt[t] <- wt[t-2] + 0.5*wt[t-1] + 2*wt[t] + 0.5*wt[t+1] + wt[t+2]
}
xt <- xt[3:202]
yt <- data.frame(xt)
yt$num <- 1:200
```



d. Obtenga y grafique la ACF y la PACF muestral de la realización obtenida en (c) y comparela con las gráficas obtenidas en (b). ¿Qué observa?



Se observa que la ACF y la PACF teoricas se comportan de igual forma que la ACF y PACF muestrales, es decir que los valores calculados de manera teorica son iguales a los calculados por la simulación para las muestrales.

2. Segundo punto.

$$X_t = 3.1 + 0.9X_{t-1} - 0.6X_{t-2} + W_t$$

Donde los W_t es un ruido blanco Gaussiano con $\mathbf{E}[W_t] = 0$ y varianza $\sigma_w^2 = 6.2$

a. Encuentre la media del proceso.

$$\begin{aligned}
\mathbf{E}[X_t] &= \mathbf{E}[3.1 + 0.9X_{t-1} - 0.6X_{t-2} + W_t] \\
\mathbf{E}[X_t] &= \mathbf{E}[3.1] + 0.9\mathbf{E}[X_{t-1}] - 0.6\mathbf{E}[X_{t-2}] + \mathbf{E}[W_t] \\
\mu &= 3.1 + 0.9\mu - 0.6\mu + 0 \\
\mu &= 3.1 + 0.3\mu \\
\mu - 0.3\mu &= 3.1 \\
\mu &= 3.1/0.7 \\
\mu &= 4.428571429
\end{aligned}$$

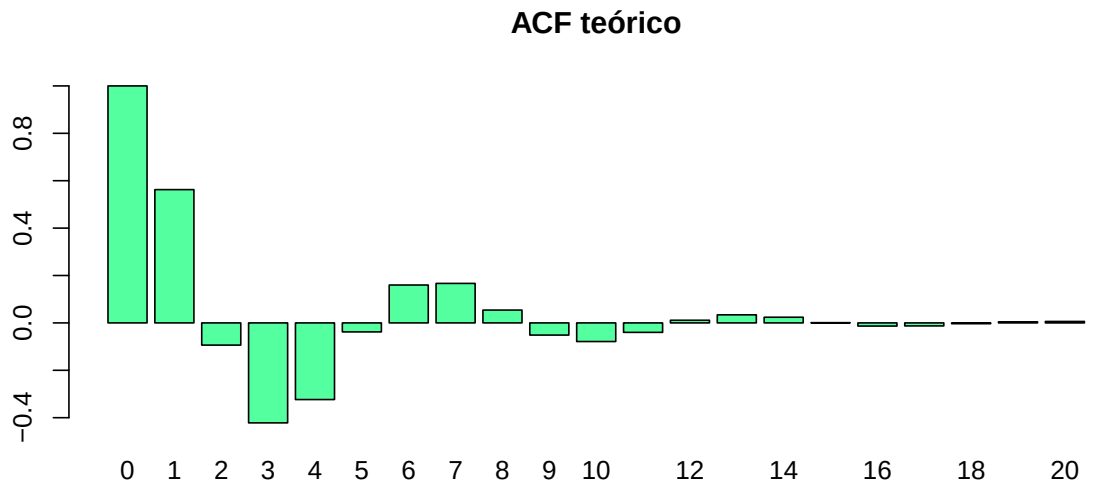
b. Encuentre la varianza del proceso.

$$\begin{aligned}
Var[X_t] &= Cov[X_t, X_t] = Cov[X_t, 3.1 + 0.9X_{t-1} - 0.6X_{t-2} + W_t] \\
&= Cov[X_t, 3.1] + 0.9Cov[X_t, X_{t-1}] - 0.6Cov[X_t, X_{t-2}] + Cov[X_t, W_t] \\
&= 0 + 0.9\gamma(1) - 0.6\gamma(2) + Cov[X_t, W_t] \\
&= 0.9 + \gamma(1) - 0.6\gamma(2) + Cov[3.1 + 0.9X_{t-1} - 0.6X_{t-2} + W_t, W_t] \\
&= 0.9\gamma(1) - 0.6\gamma(2) + Cov[W_t, W_t] \\
&= 0.9\gamma(1) - 0.6\gamma(2) + \sigma_w^2 \\
&= 0.9\gamma(1) - 0.6\gamma(2) + 6.2
\end{aligned}$$

c. Encuentre y grafique las funciones de autocorrelación (ACF) y autocorrelación parcial (PACF) para 20 lags.

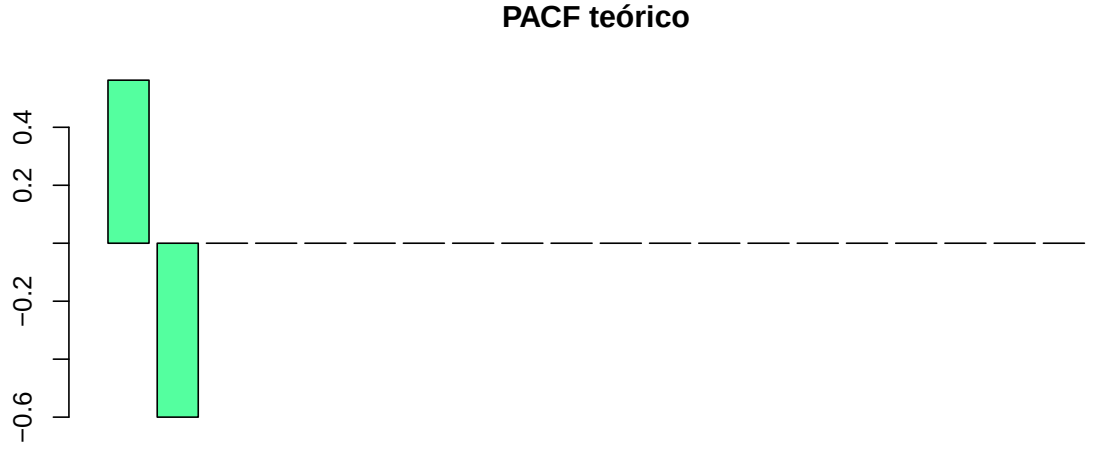
ACF

```
acf <- ARMAacf(ar = c(0.9,-0.6),lag.max = 20)
```



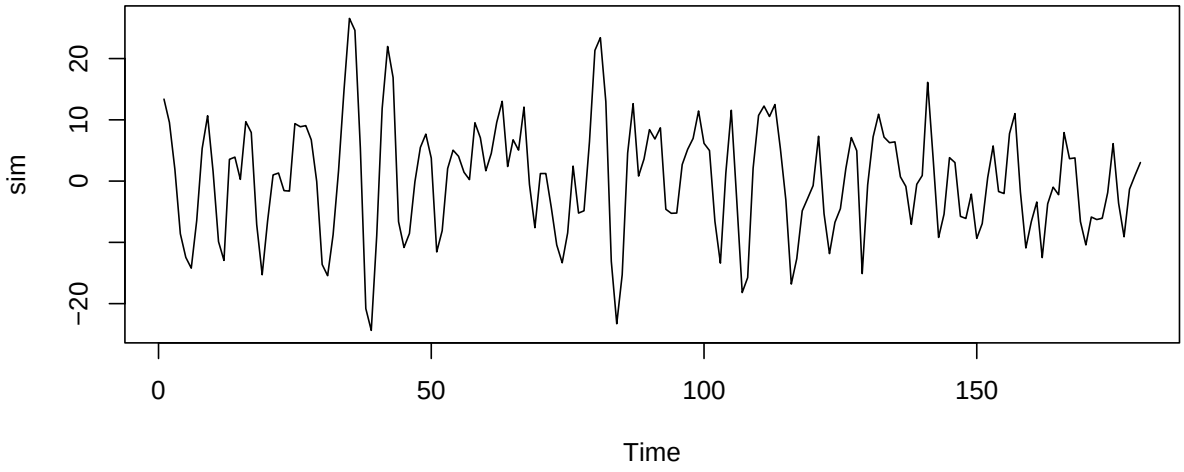
PACF

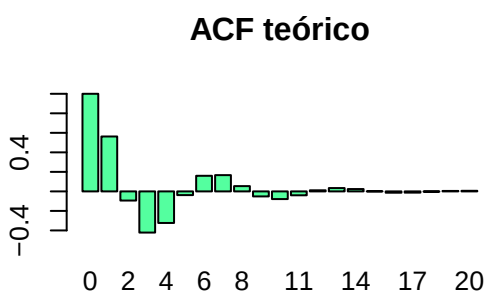
```
pacf <- ARMAacf(ar = c(0.9,-0.6),lag.max = 20,pacf = TRUE)
```

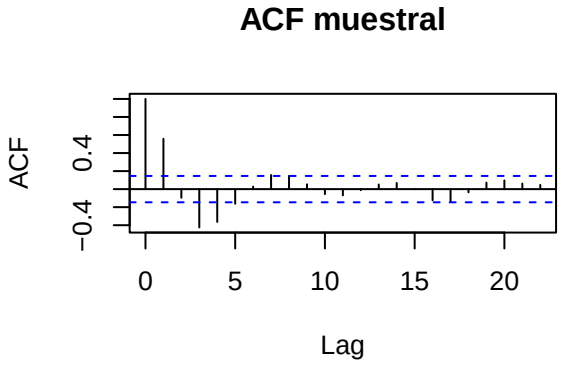


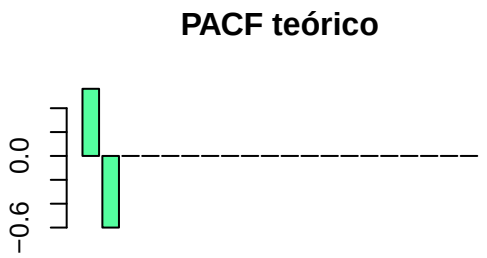
d. Simule y grafique una realización del proceso estocástico de tamaño 180. Obtenga y grafique la ACF y la PACF muestral de esta realización y comparela con las gráficas obtenidas en (c). ¿Qué observa?

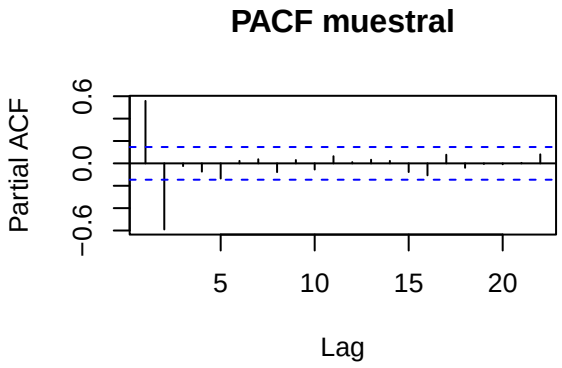
```
sim <- arima.sim(model = list(ar=c(0.9,-0.6)),n=180,sd=6.2)
```











Se observa que la ACF y la PACF teoricas se comportan de igual forma que la ACF y PACF muestrales, es decir que los valores calculados de manera teorica son iguales a los calculados por la simulación para las muestrales.

3. Tercer punto.

a. Lea cuidadosamente en RStudio las 3 bases de datos verificando que no aparezca ningún error o advertencia. ¿Cuáles son las dimensiones de cada data frame?

Cada data frame tiene respectivamente:

BD_2019 - 3563 observaciones de 23 variables

BD_2020 - 3764 observaciones de 23 variables

BD_2021 - 4122 observaciones de 23 variables

b. Una las tres bases de datos en un solo data frame. Nombrelo "datos_juntos". ¿Cuáles son las dimensiones del data frame?

Las dimensiones del data frame datos_juntos son de 11449 observaciones de 23 variables

c. Con base en el dataframe del item (b) elabore otro data frame que contenga las variables (columnas): fecha, hora, día, día de la semana (lunes, martes, . . .), semana, mes, año, número de pasajeros por hora, total de pasajeros por día, línea del metro. ¿Cuáles son las dimensiones del data frame?

Las dimensiones del data frame son de 228980 observaciones de 10 variables

d. Elabore dos data frames: uno con datos de la línea A (nombrelo "dat_lin_A") y otro con datos de la línea B (nombrelo "dat_lin_B"). Ordene los dos data frames de acuerdo a la fecha y hora. ¿Cuáles son las dimensiones de cada data frame?

Cada data frame tiene respectivamente:

dat_lin_A - 21860 observaciones de 10 variables

dat_lin_B - 21860 observaciones de 10 variables

e. Para cada línea (A y B), calcule el promedio de pasajeros por hora los lunes, los martes, los miércoles, los jueves, los viernes, los sábados y los domingos antes del 23 de marzo de 2020 (¿qué pasó en esta fecha?) y luego del 23 de marzo de 2020. ¿Qué observa? Incluya gráficos y/o tablas que ayuden a argumentar sus observaciones y comentarios.

Cuadro 1: Tabla de promedio por hora antes del 23 de marzo del 2020

hora	lunes	martes	miercoles	jueves	viernes	sabado	domingo
4	12604.78	14031.42	13019.94	13386.78	13622.46	10916.10	3129.984
5	40700.41	47065.61	44013.83	44527.81	45018.52	29231.54	10516.266
6	55943.38	64057.02	60102.39	61765.08	62941.38	37216.29	13115.203
7	48081.03	55371.75	51198.69	51654.88	52492.75	37450.24	12281.375
8	29665.80	33658.28	30925.34	33269.23	32650.81	28640.47	12395.453
9	26509.58	30619.76	27873.13	29003.50	29342.28	25232.53	14033.141

El 23 de marzo del 2020 el ministro de Salud y Protección Social, Fernando Ruiz Gómez, reiteró a los colombianos la importancia del cumplimiento del aislamiento preventivo obli-

gatorio para el control de la COVID–19 durante la emisión de un nuevo especial televisivo “Prevención y acción del coronavirus” con el expresidente de la República, Iván Duque Márquez, y algunos ministros del gabinete.

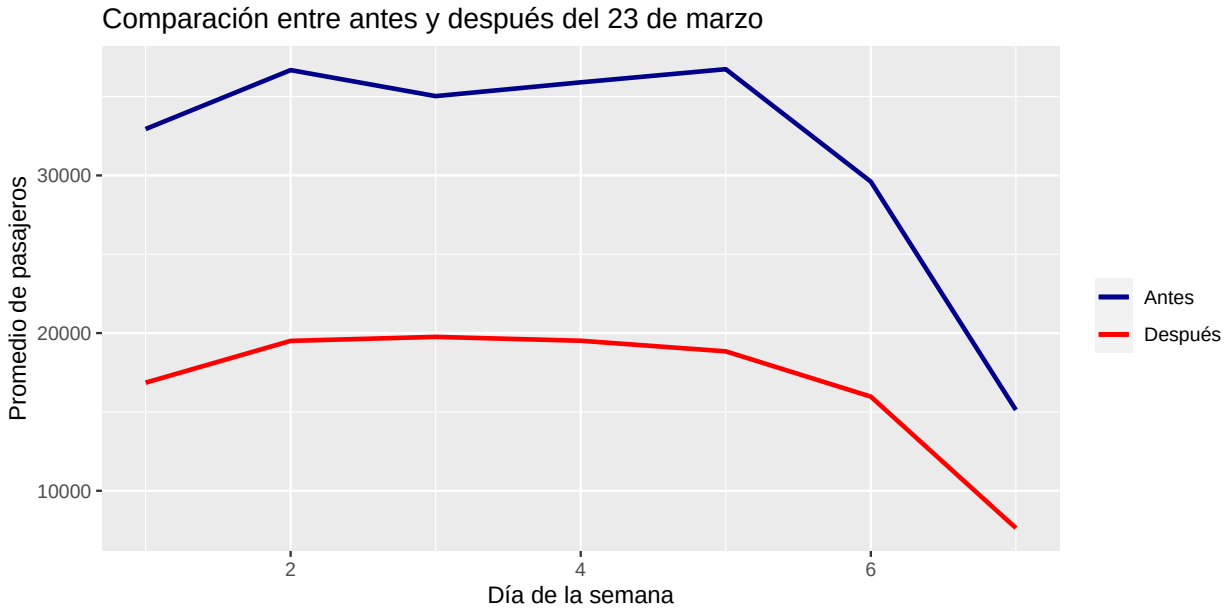
Cuadro 2: Tabla de promedio por hora despues del 23 de marzo del 2020

hora	lunes	martes	miercoles	jueves	viernes	sabado	domingo
4	9770.899	11552.04	11636.70	11566.47	10959.80	9089.965	2601.218
5	24405.461	28804.31	29361.13	29173.23	27507.50	20350.814	7264.170
6	29634.180	35512.15	36157.56	35547.36	33654.97	23334.547	8242.807
7	23220.090	27555.27	28017.88	27429.72	25766.85	19745.744	6613.614
8	15295.753	18068.30	18114.00	17707.61	16778.78	14146.651	6156.920
9	12957.483	14994.72	14990.55	14632.83	13951.49	12437.523	7136.011

A partir de las medidas tomadas por el gobierno nacional, entre esas la cuarentena, la afluencia de pasajeros del sistema de metro en la linea A y B disminuyo de manera drástica.

Cuadro 3: Tabla de promedio por día de la semana

	lunes	martes	miercoles	jueves	viernes	sabado	domingo
Antes	32940.87	36680.09	35030.10	35903.25	36737.65	29605.39	15128.478
Después	16858.69	19507.37	19762.13	19516.09	18845.03	15974.09	7641.017



Cuadro 4: Diferencia de pasajeros en porcentaje por dia después del 23 de marzo

51.17865	59.21935	59.99275	59.24584	57.20865	48.49322	23.19616

Se puede ver que hubo una diferencia de pasajeros de hasta un 59 % después de que la pandemia empezara, esto se ratifica al analizar el grafico del antes y el despues donde la linea roja que representa el despues se encuentra muy abajo comprandola con la linea azul que es antes del 23 de marzo del 2020.

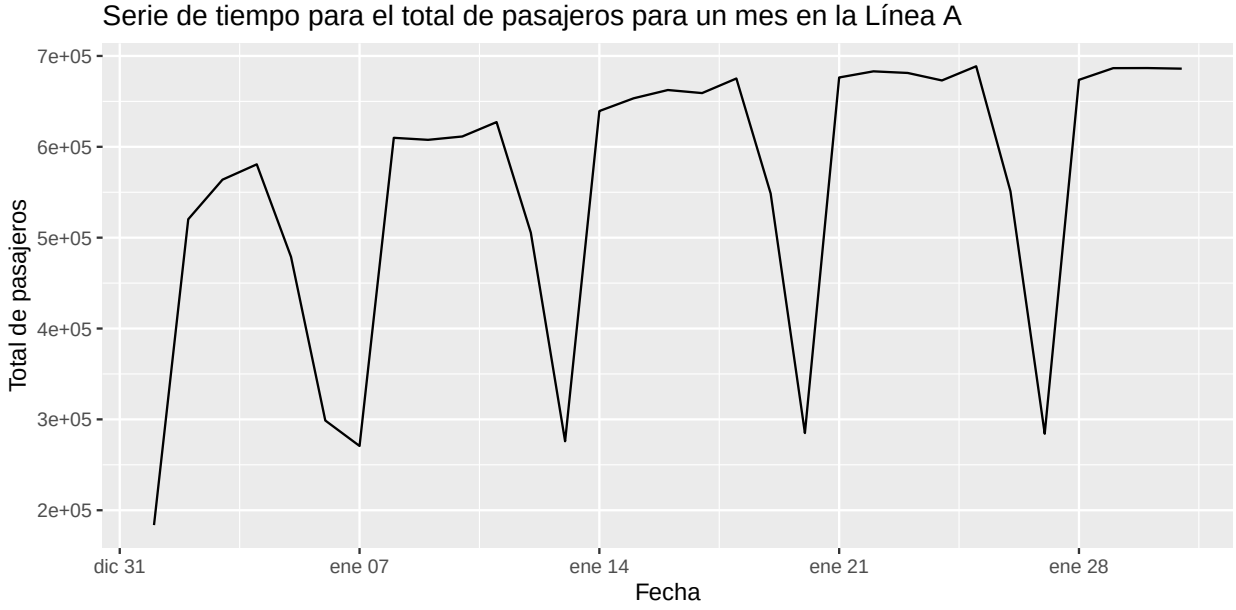
f. Obtenga dos nuevos data frames (uno para la línea A y otro para la línea B) que resuman el número total de pasajeros por día, que contengas las variables: fecha, día, día de la semana, semana, mes, año y el total de pasajeros por día.

Cada data frame tiene respectivamente:

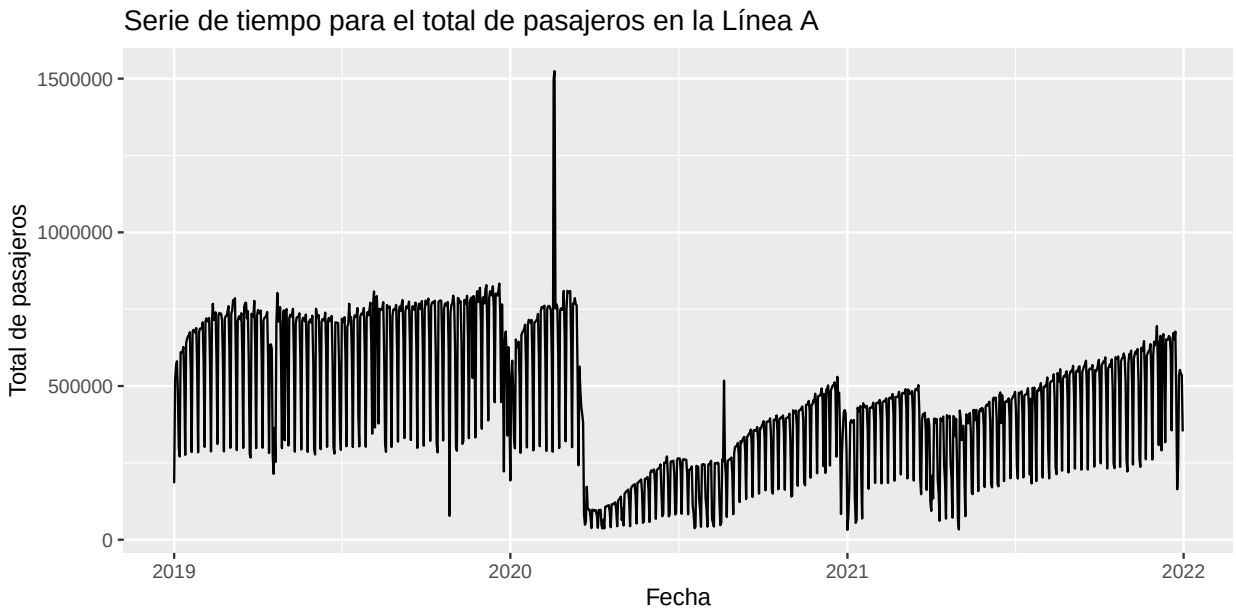
line_A - 1093 observaciones de 7 variables

linea_B - 1093 observaciones de 7 variables

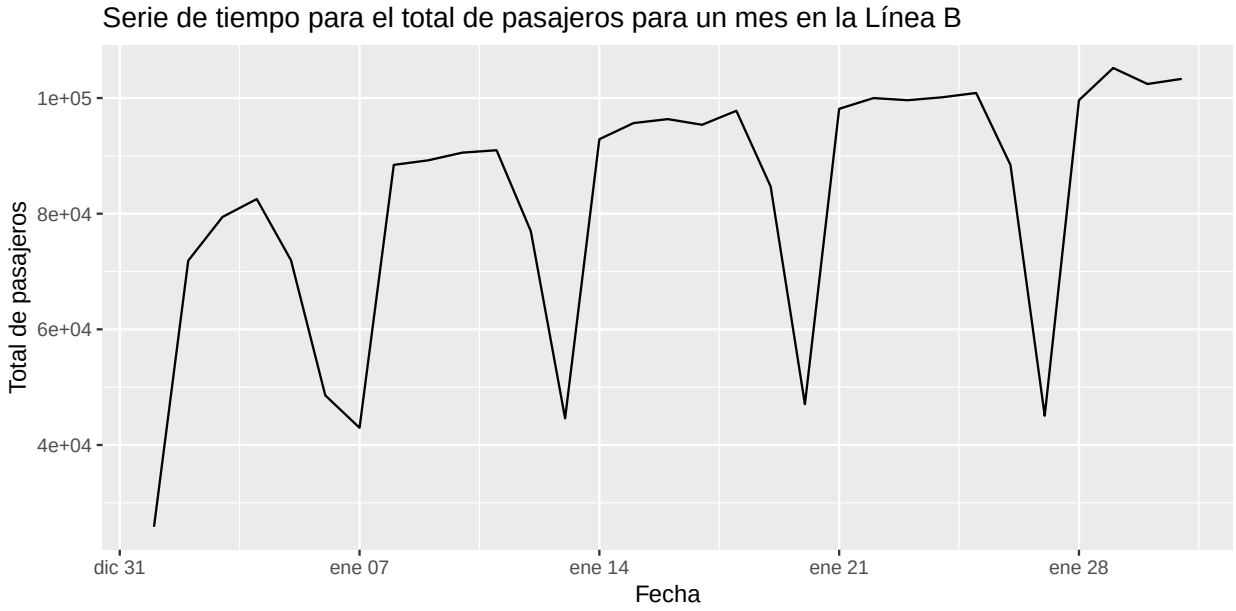
g. Para cada data frame del item (f) grafique las series del total de pasajeros por día a lo largo del tiempo. ¿Qué observa? Incluya gráficos y/o tablas que ayuden a argumentar sus observaciones y comentarios.



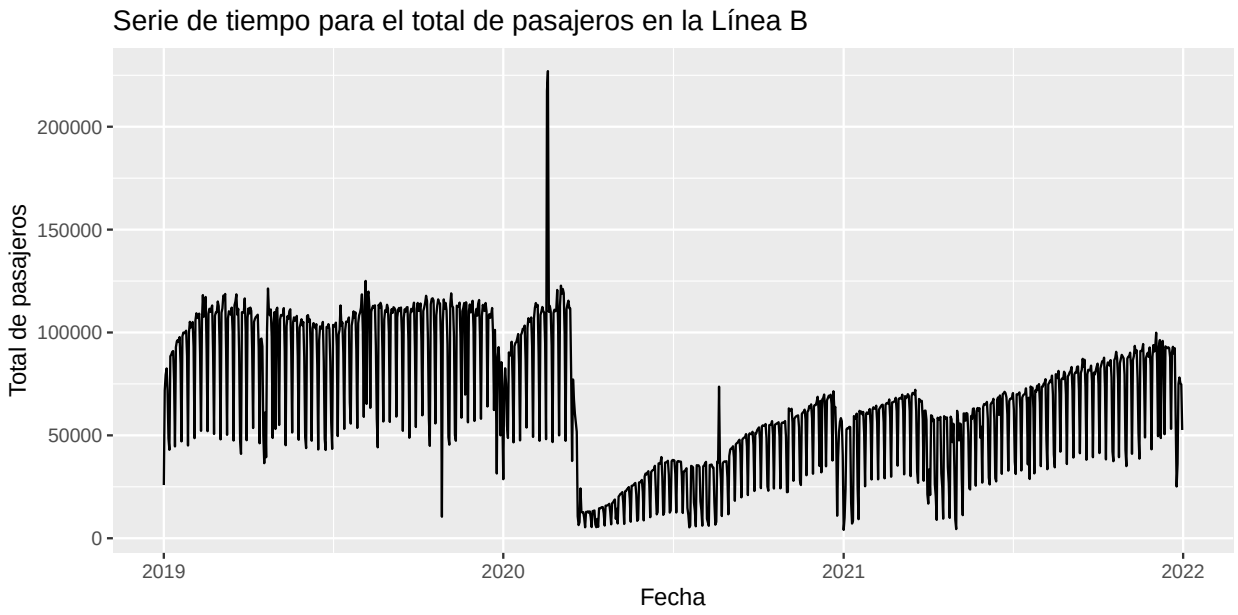
Al observar la serie de tiempo para el total de pasajeros en la linea A para un mes se puede evidenciar un comportamiento ciclico. Cada domingo se observa que la afluencia de pasajeros es menor al compararla con el resto de los dias de la semana, esto se aprecia en esos picos de caida que se observan en la serie.



Al graficar la serie de tiempo completa para el total de pasajeros en la linea A se puede apreciar que hasta antes del 23 de marzo del 2020 la serie tenia un comportamiento estacionario con una media aproximada de 500.000 pasajeros, algo interesante es que el 17 y 18 de febrero del 2020 hubo una cifra record en la cantidad de pasajeros en la linea A, ya despues del 23 de marzo del 2020, se puede apreciar que la serie tiene una tendencia creciente, esto debido al retorno a la normalidad en el sistema de metro.

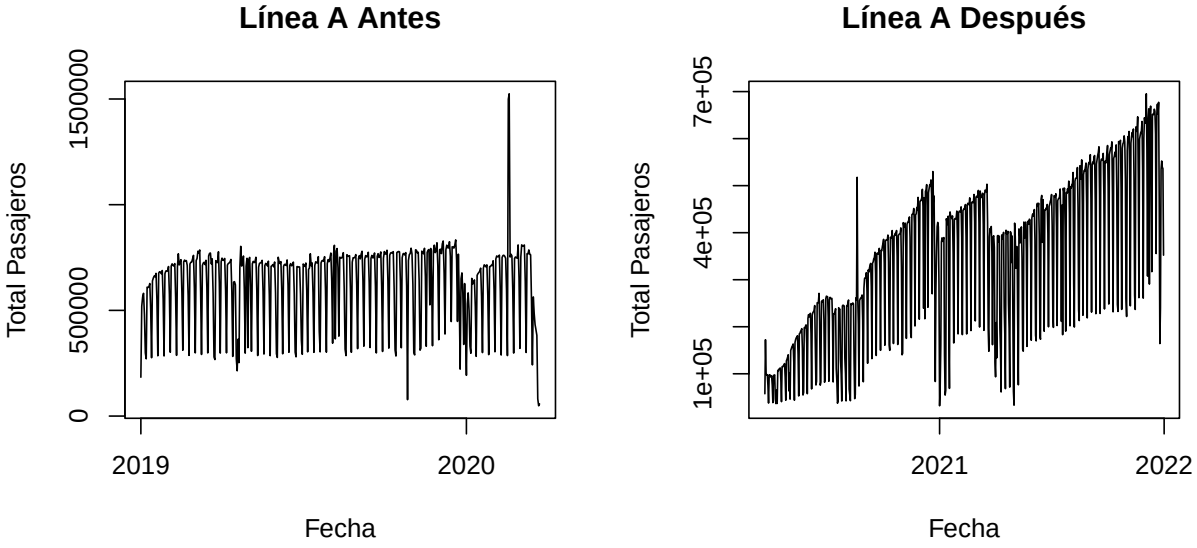


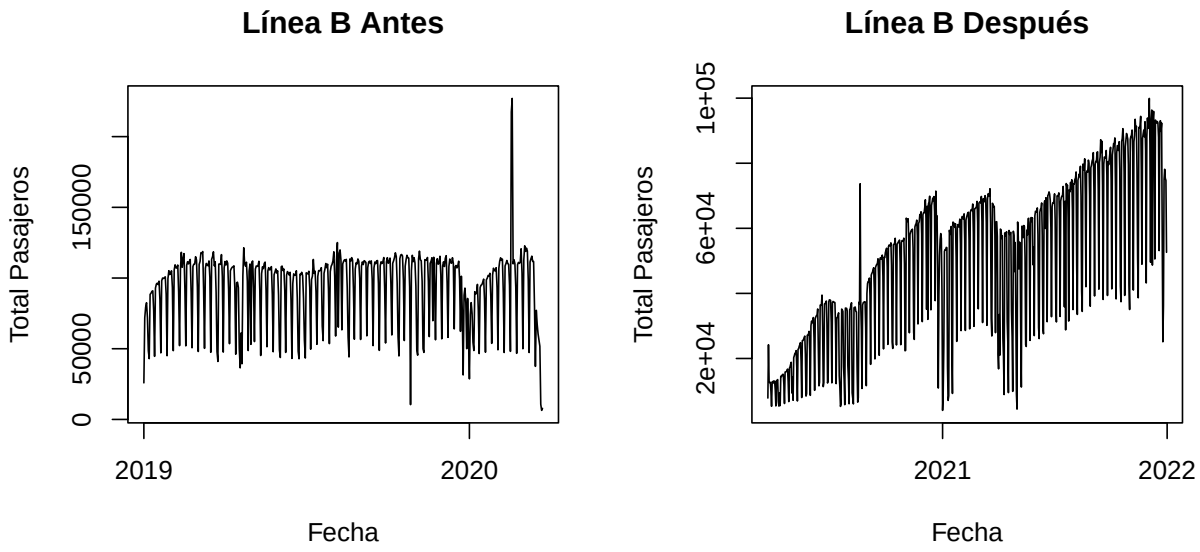
Al observar la serie de tiempo para el total de pasajeros en la linea B para un mes, se ve que su comportamiento es similar al de la linea A, presentando un comportamiento ciclico; la principal diferencia entre las dos lineas del metro, es la afluencia de pasajeros en cada una.



Al graficar la serie de tiempo completa para el total de pasajeros en la linea B se aprecia que su comportamiento es similar al de la linea A, solo que aqui la media aproximada de pasajeros es de 55000 antes del 23 de marzo, además presenta los mismos picos record en la afluencia de pasajeros y el cambio abrupto antes y despues del 23 de marzo del 2020.

h. Divida cada uno de los dos data frames obtenidos en el item (f) en antes y después del 23 de marzo de 2020. Gráfique y contaste ambos conjuntos para cada línea. ¿Observa algún comportamiento estacional o tendencia? Argumente con gráficos explicando cada uno.





Como tanto la linea A como la linea B se comportan de manera similar podemos concluir que, tanto para la linea A como para la linea B antes del 23 de marzo del 2020 presentan ambas series un comportamiento estacionario esto se ve en los ciclos que se observan en el grafico, además que cada una oscila alrededor de una media, en el caso de la linea A 500.000 y en el caso de la linea B 55000.

Ahora respecto a ambas series de tiempo pero ya observadas despues del 23 de marzo del 2020 se ve una clara tendencia lineal creciente, esto puede deberse a la normalización en el servicio de metro durante y despues de la pandemia.

i. Usando la función "lm" del RStudio, ajuste un modelo para los datos antes y otro para los datos después en cada línea que explique el número de pasajeros. Argumente por qué selecciona cada covariable y explique los resultados del "summary" de cada uno de los cuatro modelos contrastando los resultados del antes y el después en cada línea.

Modelos para linea A y B ANTES del 23 de marzo del 2020

Las covariables seleccionadas para la linea A y B fueron: dia de la semana y mes; el dia de la semana fue seleccionado ya que hay dias puntuales en que cambia de manera constante la cantidad de pasajeros en el metro, por ejemplo de lunes a viernes hay más pasajeros debido a que son dias laborales, en cambio los sabados y los domingo que son dias de descanso es menor la cantidad de personas; el mes tambien se selecciono debido a que se sabe que por ejemplo en el mes de dciembre aumentan los visitantes en la ciudad de Medellin, lo mismo en meses como abril que se encuentra la semana santa y la gente viaja a pasear por medellín, esto hace que la afluencia de personas en el servicio del metro sea mayor, cabe resaltar que todo este comportamiento fue antes de la pandemia y asi como se observo en puntos anteriores esta epoca tenia un comportamiento estacionario en la serie.

##

```
## Call:
## lm(formula = Total_Pasajeros ~ dia_semana + mes, data = linea_A_antes)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -569477  -10676   18232   46918  793581
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    560650      15104  37.119  < 2e-16 ***
## dia_semana.L   186969      14879  12.566  < 2e-16 ***
## dia_semana.Q  -294945      14886 -19.813  < 2e-16 ***
## dia_semana.C    92123      14868   6.196 1.36e-09 ***
## dia_semana^4  -113329      14863  -7.625 1.58e-13 ***
## dia_semana^5   -12216      14858  -0.822 0.411445
## dia_semana^6     9100      14860   0.612 0.540626
## mes2           121499      21821   5.568 4.54e-08 ***
## mes3            42564      22146   1.922 0.055270 .
## mes4            39187      26449   1.482 0.139176
## mes5            69672      26147   2.665 0.007998 **
## mes6            51905      26454   1.962 0.050393 .
## mes7            67300      26156   2.573 0.010415 *
## mes8            91789      26156   3.509 0.000497 ***
## mes9           115634      26454   4.371 1.55e-05 ***
## mes10          100500      26147   3.844 0.000140 ***
## mes11           96959      26449   3.666 0.000277 ***
## mes12          112964      26165   4.317 1.96e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 118900 on 430 degrees of freedom
## Multiple R-squared:  0.6184, Adjusted R-squared:  0.6033
## F-statistic: 40.99 on 17 and 430 DF,  p-value: < 2.2e-16

##
## Call:
## lm(formula = Total_Pasajeros ~ dia_semana + mes, data = linea_B_antes)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -84004  -2284   2799   6757 112896
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)     82655       2196  37.637  < 2e-16 ***
```

```
## dia_semana.L    26783       2163  12.380  < 2e-16 ***
## dia_semana.Q   -38204       2164 -17.650  < 2e-16 ***
## dia_semana.C    12486       2162   5.776 1.47e-08 ***
## dia_semana^4   -13285       2161  -6.148 1.80e-09 ***
## dia_semana^5    -2101       2160  -0.973 0.331305
## dia_semana^6     1772       2161   0.820 0.412626
## mes2            20741       3173   6.538 1.78e-10 ***
## mes3             7986       3220   2.480 0.013514 *
## mes4             8815       3846   2.292 0.022381 *
## mes5            12279       3802   3.230 0.001333 **
## mes6             6549       3846   1.703 0.089352 .
## mes7            10658       3803   2.803 0.005297 **
## mes8            17222       3803   4.529 7.70e-06 ***
## mes9            18652       3846   4.850 1.73e-06 ***
## mes10           16562       3802   4.357 1.65e-05 ***
## mes11           13494       3846   3.509 0.000497 ***
## mes12           11017       3804   2.896 0.003971 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17280 on 430 degrees of freedom
## Multiple R-squared:  0.5818, Adjusted R-squared:  0.5652
## F-statistic: 35.19 on 17 and 430 DF,  p-value: < 2.2e-16
```

Respecto al summary del modelo para la linea A y B si se obseva el valor del R^2 ajustado podemos interpretar que el 60.33 % (Linea A) y el 56.46 % (Linea B) de la variación del número de pasajeros (total de pasajeros) se explica por la variación en el porcentaje de las variables explicativas (dia de la semana, mes), el resto de la varianza puede atribuirse al azar o a otras variables que no se incorporaron en el modelo, además si se observa el p-valor de ambos modelos con una confianza del 95 % se ve que el p-valor < 0.05 siendo asi un modelo significativo.

Modelos para linea A y B DESPUES del 23 de marzo del 2020

Las covariables seleccionadas para la linea A y B fueron: fecha, dia de la semana y mes; la fecha fue seleccionada ya que hay fechas especificas donde la afluencia de personas es mayor o menor por ejemplo elecciones de votación popular en general, conciertos, partidos de futbol, etc..., pero primordialmente teniendo en cuenta que fue despues del 20 de marzo la fecha resulta ser importante ya que a medida que iba avanzando la vacunación, tratamientos e iban saliendo decretos para volver a la normalidad habian fechas especificas que influirian en la afluencia de pasajeros, como cuando se empezo a vacunar a los mayores de 18 años, o cuando fue el dia sin iba para incentivar los comercios en esa epoca de pandemia; el dia de la semana fue seleccionado ya que hay dias puntuales en que cambia de manera constante la cantidad de pasajeros en el metro, por ejemplo de lunes a viernes hay más pasajeros debido a que son dias laborales, en cambio los sabados y los domingo que son dias de descanso es menor la cantidad de personas, el mes tambien se selecciono debido a que se sabe que por

ejemplo en el mes de dciembre aumentan los visitantes en la ciudad de Medellin, lo mismo en meses como abril que se encuentra la semana santa y la gente viaja a pasear por medellín, esto hace que la afluencia de personas en el servicio del metro sea mayor.

```
##
## Call:
## lm(formula = Total_Pasajeros ~ fecha + dia_semana + mes, data = linea_A_despues)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -389116  -19643   11778   42444  221156
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -9.558e+06  3.179e+05 -30.064  < 2e-16 ***
## fecha         5.301e+02  1.704e+01  31.110  < 2e-16 ***
## dia_semana.L  1.077e+05  7.796e+03  13.808  < 2e-16 ***
## dia_semana.Q -1.739e+05  7.798e+03 -22.297  < 2e-16 ***
## dia_semana.C  5.083e+04  7.771e+03   6.542 1.26e-10 ***
## dia_semana^4 -4.033e+04  7.740e+03  -5.210 2.57e-07 ***
## dia_semana^5  3.919e+03  7.744e+03   0.506 0.612976
## dia_semana^6  3.507e+03  7.739e+03   0.453 0.650584
## mes2          7.209e+04  1.944e+04   3.709 0.000227 ***
## mes3          1.682e+04  1.785e+04   0.942 0.346447
## mes4         -8.972e+04  1.657e+04  -5.416 8.71e-08 ***
## mes5         -4.863e+04  1.643e+04  -2.960 0.003190 **
## mes6         -1.537e+04  1.655e+04  -0.929 0.353311
## mes7         -2.081e+04  1.640e+04  -1.269 0.204976
## mes8         -5.430e+03  1.645e+04  -0.330 0.741495
## mes9          4.535e+04  1.652e+04   2.745 0.006232 **
## mes10         5.435e+04  1.647e+04   3.300 0.001020 **
## mes11         6.277e+04  1.667e+04   3.765 0.000183 ***
## mes12         6.990e+04  1.660e+04   4.210 2.93e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 74530 on 627 degrees of freedom
## Multiple R-squared:  0.807,  Adjusted R-squared:  0.8015
## F-statistic: 145.7 on 18 and 627 DF,  p-value: < 2.2e-16


##
## Call:
## lm(formula = Total_Pasajeros ~ fecha + dia_semana + mes, data = linea_B_despues)
##
```

```
## Residuals:
##    Min     1Q Median     3Q    Max
## -55183  -2493   1818   5470  31341
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)  -1.518e+06  4.466e+04 -34.002  < 2e-16 ***
## fecha         8.388e+01  2.393e+00  35.047  < 2e-16 ***
## dia_semana.L  1.457e+04  1.095e+03  13.302  < 2e-16 ***
## dia_semana.Q -2.351e+04  1.095e+03 -21.459  < 2e-16 ***
## dia_semana.C  6.954e+03  1.092e+03   6.371 3.63e-10 ***
## dia_semana^4 -4.738e+03  1.087e+03  -4.358 1.53e-05 ***
## dia_semana^5  7.706e+02  1.088e+03   0.708 0.478937
## dia_semana^6  6.515e+02  1.087e+03   0.599 0.549178
## mes2          1.133e+04  2.731e+03   4.151 3.78e-05 ***
## mes3          5.654e+03  2.507e+03   2.255 0.024479 *
## mes4         -1.084e+04  2.327e+03  -4.658 3.90e-06 ***
## mes5         -5.081e+03  2.308e+03  -2.202 0.028051 *
## mes6          6.407e+02  2.325e+03   0.276 0.782937
## mes7         -5.470e+02  2.303e+03  -0.237 0.812377
## mes8          1.590e+03  2.311e+03   0.688 0.491754
## mes9          8.707e+03  2.321e+03   3.751 0.000192 ***
## mes10         9.896e+03  2.313e+03   4.278 2.18e-05 ***
## mes11         1.100e+04  2.342e+03   4.695 3.27e-06 ***
## mes12         9.026e+03  2.332e+03   3.870 0.000120 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 10470 on 627 degrees of freedom
## Multiple R-squared:  0.821,  Adjusted R-squared:  0.8159
## F-statistic: 159.8 on 18 and 627 DF,  p-value: < 2.2e-16
```

Respecto al summary del modelo para la linea A y B si se obseva el valor del R^2 ajustado podemos interpretar que el 80.15 % (Linea A) y el 81.59 % (Linea B) de la variación del número de pasajeros (total de pasajeros) se explica por la variación en el porcentaje de las variables explicativas (fecha, dia de la semana, mes), el resto de la varianza puede atribuirse al azar o a otras variables que no se incorporaron en el modelo; además si se observa el p-valor de ambos modelos con una confianza del 95 % se ve que el p-valor < 0.05 siendo asi un modelo significativo.

4. Anexos.

En el siguiente link se encuentra todo el trabajo y los codigos empleados para su solución: