

Universidad Nacional De Colombia
Sede Medellín



Facultad de Ciencias
Departamento de Estadística

Trabajo 1
Procesos estacionarios

Juan David García Zapata
Juan Camilo Miranda Paz

Series de tiempo Univariadas

Marzo, 2024

Índice

1. Punto 1	2
1.1. a)	2
1.2. b)	3
1.3. c)	5
2. Punto 2	7
2.1. a)	7
2.2. b)	9
3. Punto 3	10
3.1. a)	10
3.2. b)	13
3.3. c)	15
4. Punto 4	16
4.1. a)	16
4.2. b)	19
4.3. c)	19

1. Punto 1

Enunciado

Considere el proceso Z_t , definido por:

$$Z_t = A_1 \cos(\omega t) + A_2 \sin(\omega t),$$

donde ω corresponde a un parámetro de frecuencia constante. A_1 y A_2 son 2 variables normales independientes de media cero y varianza σ^2 .

- a) Investigue, consulte las características de este proceso, su media, su varianza y su función de auto-correlación (no es necesario hacer o reportar los cálculos).
- b) ¿Es el proceso Z_t estacionario?.
- c) Grafique 2 (dos) realizaciones de este proceso.

1.1. a)

Media: Ya conocemos que A_1 y A_2 son 2 variables normales **independientes** de media 0 y varianza σ^2 es decir:

$$E[A_1] = 0$$

y

$$E[A_2] = 0$$

Entonces:

$$\begin{aligned} E[Z_t] &= E[A_1 \cos(\omega t) + A_2 \sin(\omega t)] \\ &= E[0 * \cos(\omega t) + 0 * \sin(\omega t)] \\ &= 0 \end{aligned}$$

Varianza: La varianza de Z_t puede ser hallada utilizando la propiedad de la independencia de A_1 y A_2 . La varianza de la suma es la suma de las varianzas:

$$\begin{aligned} \text{Var}(Z_t) &= \text{Var}(A_1 \cos(\omega t)) + \text{Var}(A_2 \sin(\omega t)) \\ &= \cos^2(\omega t) \text{Var}(A_1) + \sin^2(\omega t) \text{Var}(A_2) \\ &= \cos^2(\omega t) \sigma^2 + \sin^2(\omega t) \sigma^2 \\ &= \sigma^2 (\cos^2(\omega t) + \sin^2(\omega t)) \\ &= \sigma^2 \cdot 1 \\ &= \sigma^2 \end{aligned}$$

Demostrado que la varianza de Z_t es constante y no depende del tiempo t .

Auto-correlación: La función de autocorrelación $R(h)$ para este proceso se define como:

$$R(h) = E[Z_t Z_{t+h}]$$

Desarrollando el producto y utilizando la independencia de A_1 y A_2 , tenemos:

$$\begin{aligned} R(h) &= E[(A_1 \cos(\omega t) + A_2 \sin(\omega t))(A_1 \cos(\omega(t+h)) + A_2 \sin(\omega(t+h)))] \\ &= E[A_1^2 \cos(\omega t) \cos(\omega(t+h))] + E[A_2^2 \sin(\omega t) \sin(\omega(t+h))] \\ &\quad + E[A_1 A_2 \cos(\omega t) \sin(\omega(t+h))] + E[A_2 A_1 \sin(\omega t) \cos(\omega(t+h))] \end{aligned}$$

Dado que A_1 y A_2 son independientes y tienen una varianza igual a σ^2 y $\cos(\omega t)$ y $\sin(\omega t)$ son ortogonales entre si, la función de autocorrelación se reduce a:

$$\sigma^2 \cos(\omega h)$$

Así, la función de autocorrelación para este proceso armónico (o sinusoidal) es una función coseno con amplitud σ^2 y frecuencia ω .

Con esto podemos concluir las características principales del proceso Z_t , con media cero, varianza σ^2 y función de autocorrelación dada por $\sigma^2 \cos(\omega h)$.

1.2. b)

Teóricamente el proceso Z_t es estacionario en sentido débil, es un proceso estacionario débil es aquel en las cuales las propiedades estadísticas del proceso no cambien con el tiempo. Esto significa que las propiedades como la media y la varianza y la autocorrelación son constantes a lo largo del tiempo o dependen solo de la separación temporal.

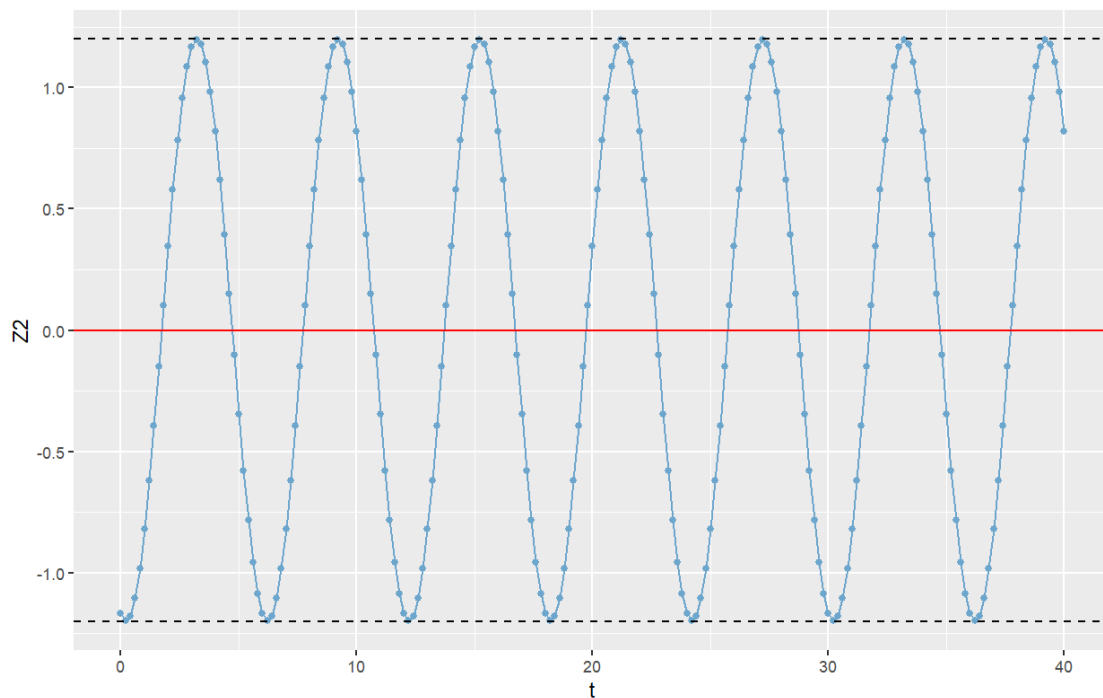
Para demostrar que Z_t es estacionario en sentido débil, podemos verificar que las propiedades estadísticas como la media, la varianza y la función de autocorrelación no cambian con el tiempo:

- **Media:** el el ítem anterior se demostró que la media del proceso(Z_t) es cero, lo cual es constante para todo t .
- **Varianza:** De igual forma, en el ítem anterior se demostró que la varianza de Z_t es σ^2 que es constante independientemente del tiempo
- **Función de autocorrelación.** es igual a $R(h) = \sigma^2 \cos(\omega h)$ que depende de la separación temporal h y no del tiempo absoluto t , por lo tanto la autocorrelación también es constante a lo largo del tiempo.

Dado estas propiedades no dependen del tiempo t , podemos concluir que el proceso Z_t es estacionario en sentido débil.

Verificación: Gráficamente podemos verificarlo de la siguiente forma:

```
#Cargar librerías
library(ggplot2)
# Realizaciones del proceso
set.seed(1001660790)
# Parámetros del proceso
w <- pi/3
s <- 1
A1_2 <- rnorm(1)
A2_2 <- rnorm(1)
t <- seq(0, 40, 0.2)
Z2 <- A1_2 * cos(w * t) + A2_2 * sin(w * t)
# Calcular la media y sigma^2
media <- rep(0, length(t))
varianza <- rep(s^2, length(t))
# Gráfico
ggplot(data.frame(t, Z2, media, varianza)) +
  geom_point(aes(x = t, y = Z2), color = "skyblue3") +
  geom_line(aes(x=t, y=Z2),color = "skyblue3") +
  geom_hline(yintercept = 0, color = "red") +
  geom_hline(yintercept = c(-1.2, 1.2), linetype = "dashed",
    color = "black")
```



Al examinar con atención la imagen anterior, resulta evidente que se satisfacen todas las propiedades estadísticas previamente establecidas, lo cual respalda la validez de nuestros análisis.

1.3. c)

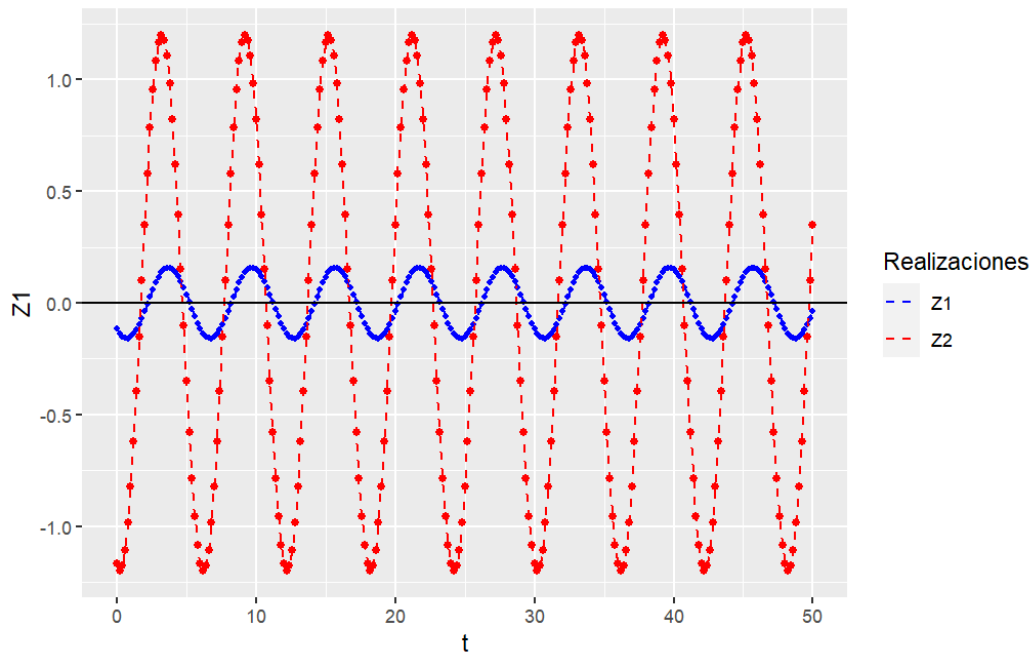
Gráfico de 2 realizaciones del proceso Z_t donde tendremos en cuenta $t=1,2,3,4,\dots,50$

```
set.seed(1001660790)
  # Parámetros del proceso
w <- pi/3
s <- 1
A1_1 <- rnorm(1) # para A1 de la realización 1
A2_1 <- rnorm(1) # para A2 de la realización 1
A1_2 <- rnorm(1) # para A1 de la realización 2
A2_2 <- rnorm(1) # para A2 de la realización 2

# Realizaciones del proceso
t <- seq(0, 50, 0.2)
Z1 <- A1_1 * cos(w * t) + A2_1 * sin(w * t)
Z2 <- A1_2 * cos(w * t) + A2_2 * sin(w * t)

# Gráfico

ggplot(data.frame(t, Z1, Z2)) +
  geom_point(aes(x = t, y = Z1), color = "red", size=1.5) +
  geom_point(aes(x = t, y = Z2), color = "blue", size=1) +
  geom_line(aes(x = t, y = Z1, color = "red"), lty = 2) +
  geom_line(aes(x = t, y = Z2, color = "blue"), lty = 2) +
  geom_hline(yintercept = 0, color = "black") +
  scale_color_identity(guide = 'legend', labels = c("Z1", "Z2")) +
  guides(color = guide_legend(title = "Realizaciones"))
```



En la representación gráfica del proceso, es evidente que su comportamiento está determinado únicamente por dos factores: la frecuencia ω y la amplitud σ^2 .

- ω , la frecuencia, dicta la rapidez con la que se producen las oscilaciones en el proceso. Un valor mayor de ω conlleva oscilaciones más rápidas, mientras que un valor menor resulta en oscilaciones más lentas o incluso en ausencia de estas cuando $\omega = 0$
- Por otro lado, σ^2 , la amplitud, establece la magnitud de estas oscilaciones. A medida que σ^2 aumenta, las oscilaciones se vuelven más pronunciadas, mientras que una disminución en σ^2 reduce su intensidad

Por lo tanto, al observar la gráfica del proceso, se pueden identificar claramente los efectos de ω y σ^2 en las 2 realizaciones del proceso, los cuales son los parámetros fundamentales que determinan su comportamiento.

2. Punto 2

Enunciado

El siguiente código en R genera una serie temporal con 100 datos. Usted va a identificar el tipo de proceso subyacente, realizando una gráfica de los datos y también graficando los correlogramas:

```
library(ggplot2)
library(dplyr)
library(hrbrthemes)

generador <- function(cedula)
{ set.seed(cedula)
  data <- rnorm(100)
  data
}
times <- seq(1,100)
values <- generador( cedula)
Datos = data.frame(times,values)
# Most basic bubble plot
p <- ggplot(Datos, aes(x=times, y=values)) +
  geom_line( color="steelblue") +
  geom_point() +
  xlab("") +
  theme\_ipsum() +
  theme(axis.text.x=element\_text(angle=60, hjust=1))
p
```

- Ejecute el código anterior, colocando en el campo de la cédula, el número de cédula de uno de los integrantes del grupo.
- Grafique la función de autocorrelación ACF y la función de autocorrelación parcial PACF estimadas de los datos anteriores. Identifique el proceso.

2.1. a)

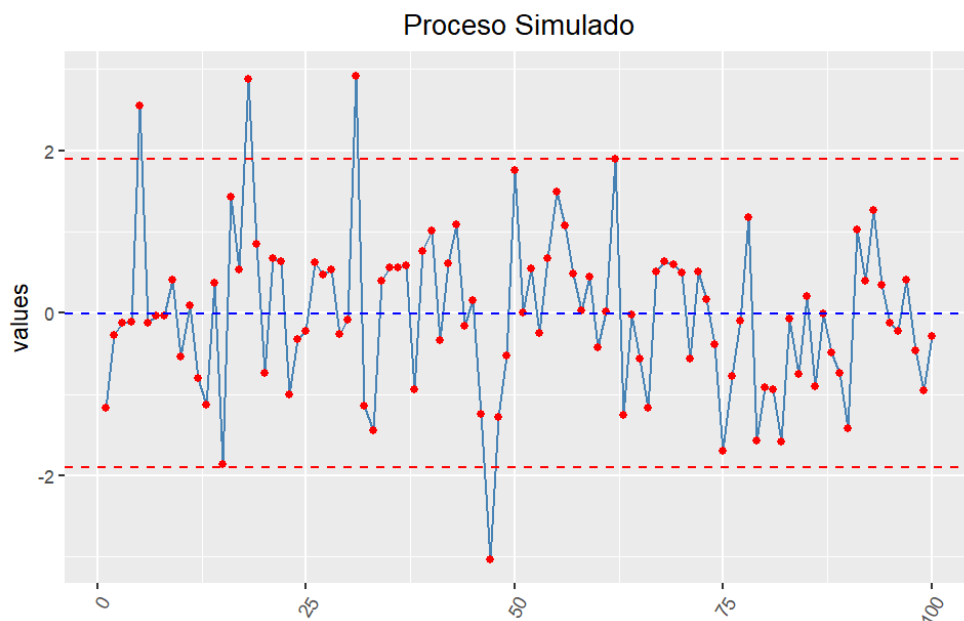
Utilizando el código proporcionado en el enunciado, pero haciéndole algunas mejoras, tenemos el siguiente resultado utilizando la cédula **10011660790** del estudiante Juan David, así:


```

library(ggplot2)
library(dplyr)
library(hrbrthemes)
generador <- function(cedula){
  set.seed(cedula)
  data <- rnorm(100)
  data
}
times <- seq(1,100)
values <- generador(1001660790)
Datos = data.frame(times,values)
# Most basic bubble plot
p <- ggplot(Datos, aes(x=times, y=values)) +
  geom_line( color="steelblue") +
  geom_point(color = "red") +
  xlab("") +
  geom_hline(yintercept=1.9, color = "red", linetype = "dashed") +
  ggtitle("Proceso Simulado") +
  theme(plot.title = element_text(hjust = 0.5))+
  geom_hline(yintercept = mean(Datos$values), color = "blue",
  linetype = "dashed")+xlab("") +
  geom_hline(yintercept=-1.9, color = "red", linetype = "dashed") +
  theme(axis.text.x=element_text(angle=60, hjust=1))
p

```

Obtenemos:

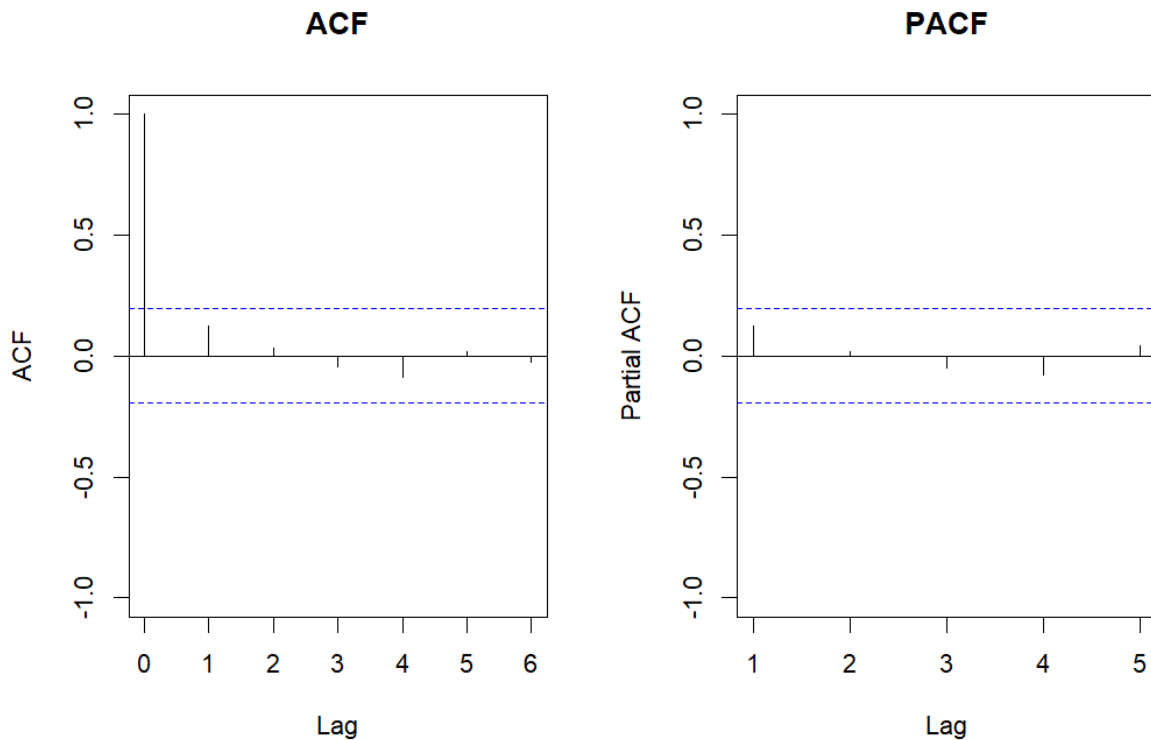


Se observa que los datos revelan una media constante de 0, lo cual es indicativo de un proceso de ruido blanco. Sin embargo, al examinar más de cerca el intervalo $(-1.9, 1.9)$, podemos ver hay algunas constrastes de menor y mayor variabilidad.

2.2. b)

Graficamos la función ACF y PACF estimadas, así:

```
par(mfrow=c(1,2))
acf(Datos$values, ylim=c(-1,1), main= "ACF", lag.max = 6)
pacf(Datos$values, ylim=c(-1,1), main= "PACF", lag.max = 5)
```



De acuerdo con los gráficos anteriores, se evidencia un rezago en cero de la ACF, además de la estacionariedad presentada en el punto anterior. Por lo tanto, un modelo de ruido blanco podría ser adecuado para describir los datos.

3. Punto 3

Enunciado

- a) Suponiendo que los datos del punto anterior corresponde a una realización de un proceso de un **ruido blanco** a_t , usted va a construir 2 realizaciones de un proceso de media móvil MA(1), **invertibles**, una correlación $\rho > 0$ y otra con $-\rho$. **Grafique** las correspondientes realizaciones o trayectorias y **escriba las ecuaciones que los especifican**. ¿Qué diferencias observa?
- b) Para cada una de las trayectorias del proceso de MA(1) arriba, grafique la ACF y la PACF
- c) Con los datos correspondientes a la realización del proceso MA(2) en el 1er ítem, **grafique** diagramas de dispersión de Z_t vs Z_{t-1} vs Z_{t-2} .

3.1. a)

Se define el proceso Z_t como:

$$Z_t = \mu + a_t - \theta a_{t-1}$$

donde

$$Z_t = \mu + \theta(B)a_t,$$

y $\theta(B) = 1 - \theta B$, y $\{a_t\}$ es un proceso de ruido blanco de media cero, esto implica que, para que el proceso sea invertible es necesario que las raíces de $\theta(B) = 1 - \theta B = 0$ caigan fuera del círculo unidad. Esto se cumple si $|\theta| < 1$.

La auto-covarianza γ_k para un proceso MA(1) están dadas por:

$$\gamma_k = \begin{cases} (1 + \theta^2)\sigma^2 & \text{si } k = 0 \\ -\theta\sigma^2 & \text{si } k = 1 \\ 0 & \text{si } k > 1 \end{cases}$$

y la función de autocorrelación ρ_k por:

$$\rho_k = \begin{cases} -\frac{\theta}{1+\theta^2} & \text{si } k = 1 \\ 0 & \text{si } k > 1 \end{cases}$$

Por tanto, la ACF se anula (presenta un corte) a partir de $k = 2$.

Teniendo en cuenta esto y la condición dada por el problema, encontramos las raíces del polinomio

$$\theta_1^2 \rho - \theta + \rho = 0$$

y tomamos la que cumpla la primera condición de que $\rho > 0$, con $\rho = 0.2$. Se calcula θ_1 utilizando la fórmula cuadrática:

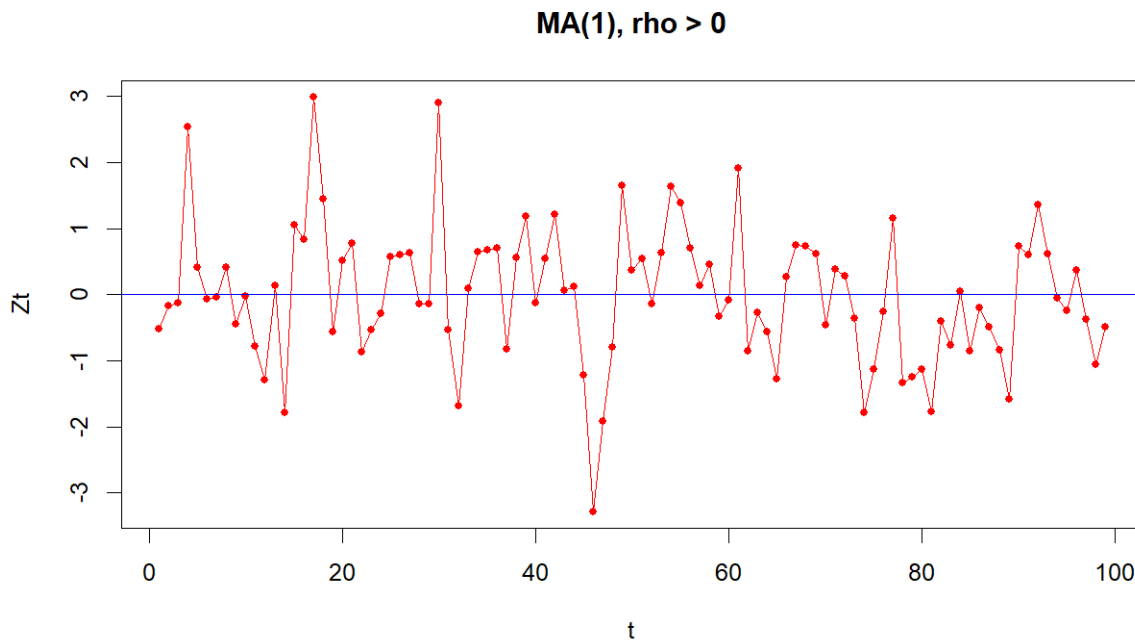
$$\theta_1 = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{-(-1) \pm \sqrt{(-1)^2 - 4(0.2)(0.2)}}{2(0.2)}$$

Obteniendo los siguientes resultados:

$$\theta_1 = 4.791 \quad \text{ó} \quad \theta_1 = 0.209$$

Tomamos el valor de $\theta_1 = 0.209$, que cumple la condición de que $|\theta_1| < 1$:

```
theta1 <- 0.209 # Raiz
a <- Datos$values
ma_1 <- c(NA, sapply(2:length(a), function(t) theta1 * a[t-1] + a[t]))
plot(unlist(ma_1), type = "o", col="red", main="MA(1), rho > 0",
     pch = 20, xlab = "t",
     ylab = "Zt")
abline(h=0, col="blue")
```



La cual tiene la ecuación:

$$\theta_1^2(0.2) - \theta_1 + (0.2) = 0$$

Para el proceso MA(1) con $-\rho$, repetimos el proceso pero con $\rho = -0.2$, así:

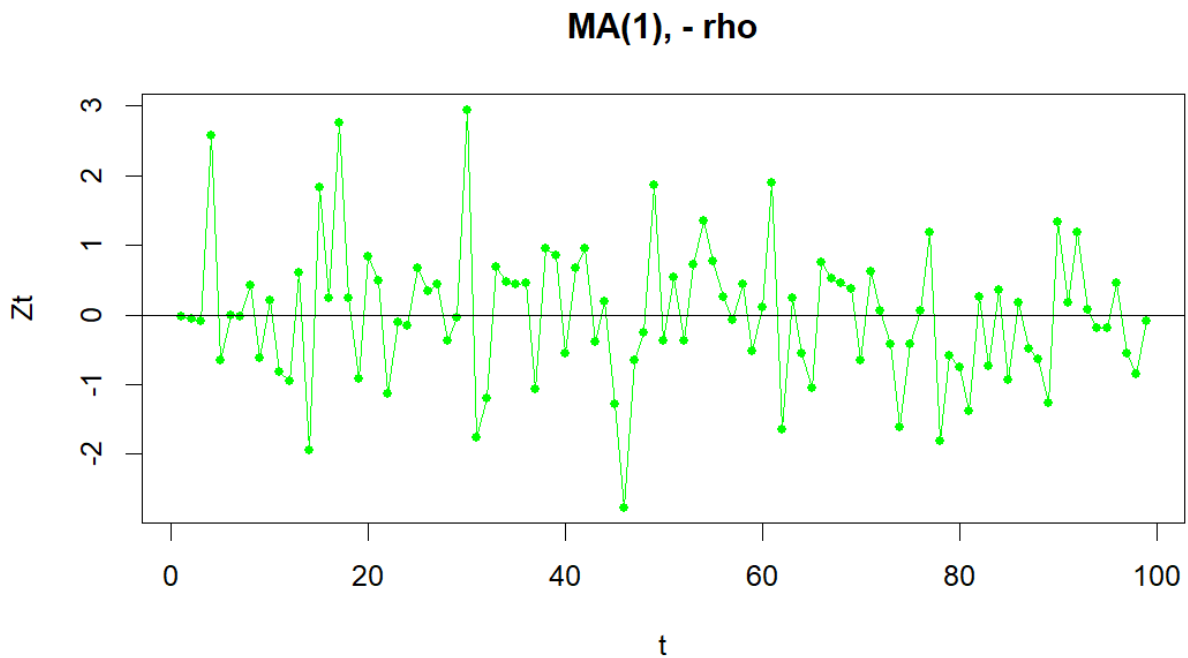
$$\theta_2 = \frac{-(-1) \pm \sqrt{(-1)^2 - 4(-0.2)(-0.2)}}{2(-0.2)}$$

Obteniendo los siguientes resultados:

$$\theta_2 = -4.791 \quad \text{ó} \quad \theta_1 = -0.209$$

Tomamos el valor de $\theta_2 = -0.209$, que cumple la condición de que $|\theta_2| < 1$:

```
theta2 <- -0.209 # Raíz
a <- Datos$values
ma_2 <- c(NA, sapply(2:length(a), function(t) theta2 * a[t-1] + a[t]))
plot(unlist(ma_2), type = "o", col="green", main="MA(1), - rho",
     pch = 20, xlab = "t", ylab = "Zt")
abline(h=0, col="black")
```

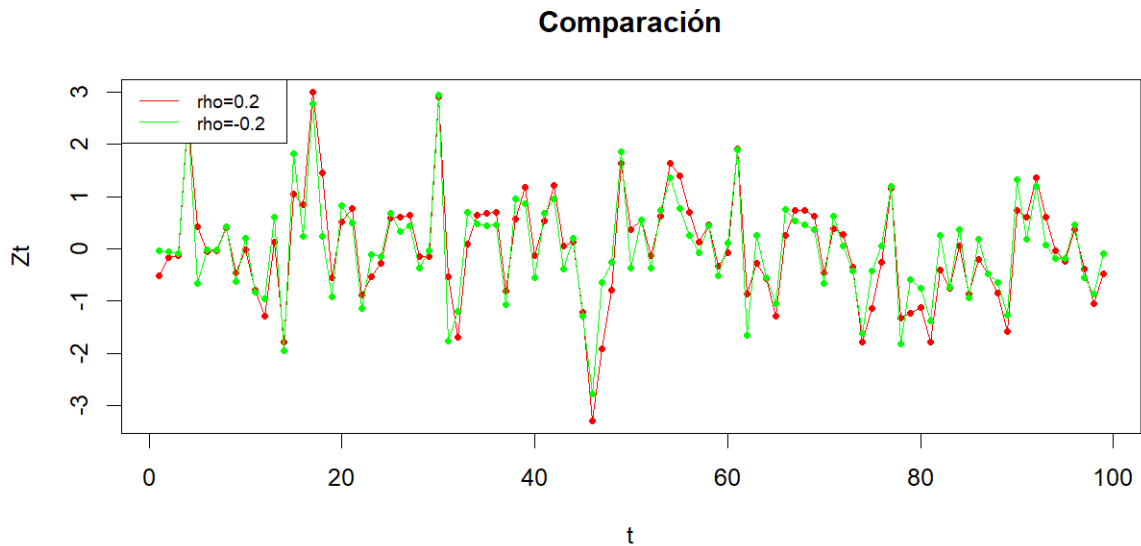


Con la ecuación:

$$-\theta_2^2(0.2) - \theta_2 - (0.2) = 0$$

Comparamos los procesos:

```
plot(unlist(ma_1), type = "o", col = "red",
     main="Comparación", xlab = "t", ylab= "Zt", pch = 20)
lines(unlist(ma_2), type = "o", col="green", pch = 20)
legend("topleft", legend = c("rho=0.2", "rho=-0.2"), lty=1,
      col=c("red","green") , cex = 0.75)
```

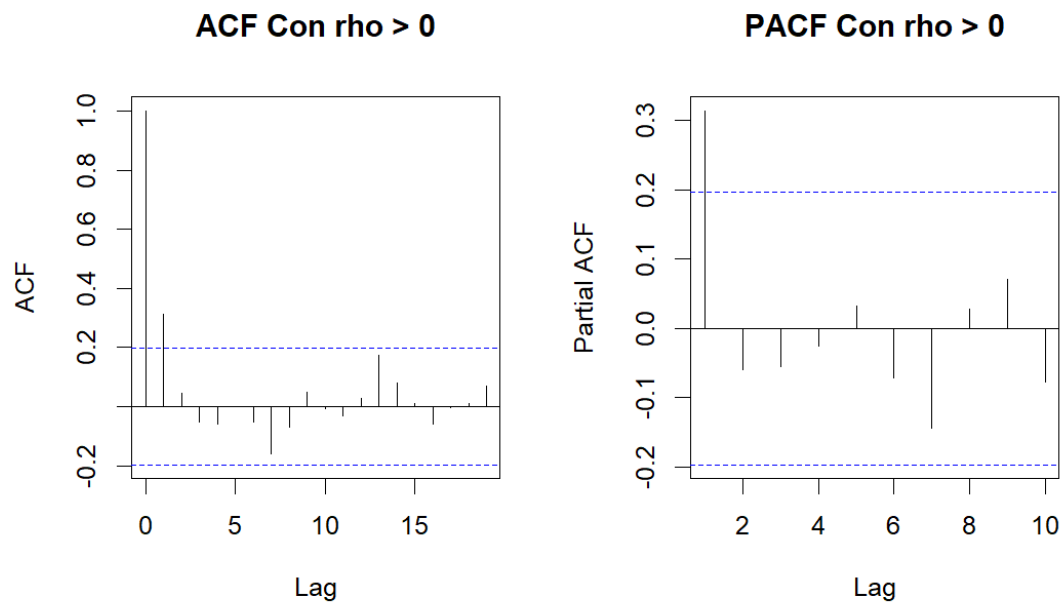


Hemos observado que el comportamiento del proceso MA(1) cuando ρ es positivo tiende a superponerse al comportamiento observado cuando ρ es negativo. Este fenómeno se atribuye al efecto del coeficiente negativo en la media móvil, el cual induce a que la correlación de las tendencias positivas a lo largo de la serie temporal prevalezca sobre las tendencias negativas. Esto se manifiesta claramente en la comparación de las curvas de la serie, donde la correspondiente a tendencias positivas alcanza picos más altos en comparación con su contraparte. De manera similar, este mismo efecto hace que, en los puntos de menor elevación, la correlación negativa se posicione consistentemente por debajo a lo largo de la serie temporal.

3.2. b)

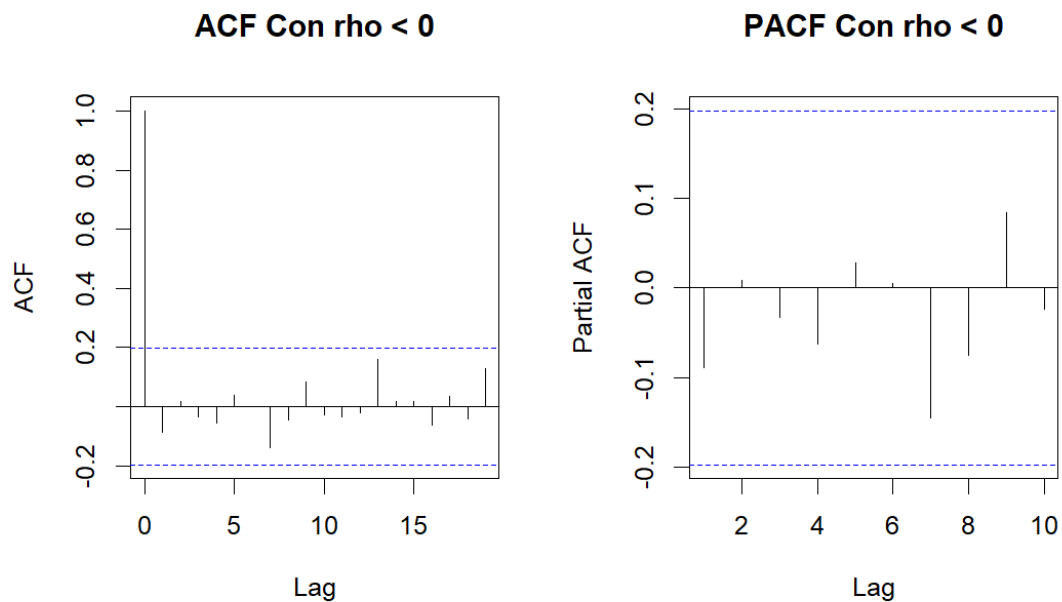
Para cada uno de los procesos MA(1), graficamos ACF y PACF:

```
par(mfrow= c(1,2))
acf(unlist(ma_1), main= "ACF Con rho > 0")
pacf(unlist(ma_1), lag.max = 10, main= "PACF Con rho > 0")
```



Podemos observar que gracias a la forma de la ACF con $\rho > 0$, podemos indicar que se trata de una serie de tiempo estacional

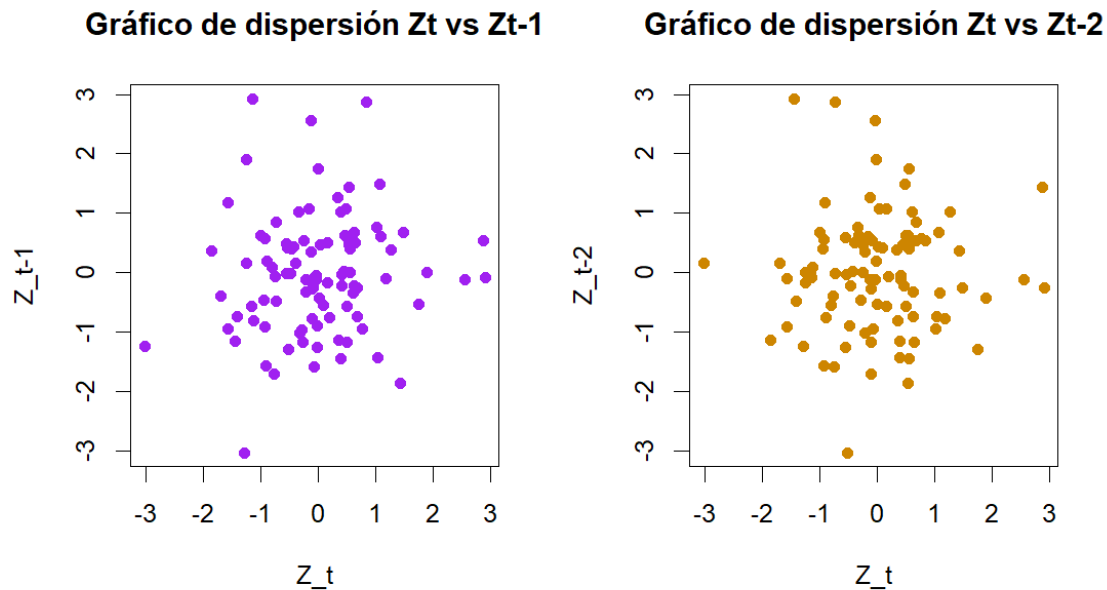
```
par(mfrow= c(1,2))
acf(unlist(ma_2), main= "ACF Con rho < 0")
pacf(unlist(ma_2), lag.max = 10, main= "PACF Con rho < 0")
```



3.3. c)

Diagramas de dispersión de Z_t vs Z_{t-1} y Z_t vs Z_{t-2} :

```
par(mfrow=c(1,2))
plot(Datos$values[2:100], Datos$values[1:99],
     xlab = "Z_t", ylab = "Z_t-1", col="purple", pch=16,
     main= "Gráfico de dispersión Zt vs Zt-1")
plot(Datos$values[3:100], Datos$values[1:98],
     xlab = "Z_t", ylab = "Z_t-2", col="orange3", pch=16,
     main= "Gráfico de dispersión Zt vs Zt-2")
```

Los diagramas de dispersión indican que no hay una correlación clara entre los valores de Z_t vs Z_{t-1} y vs Z_{t-2}

4. Punto 4

Enunciado

- a) De la misma manera que el punto anterior considere que la realización del 1er punto corresponde a un proceso de **ruido blanco** a_t . Esta vez construya la correspondiente realización de un proceso de **autoregresivo** AR(2), que sea **estacionario** con una estructura de correlación **cíclica**. **Grafique** la correspondiente realización o trayectoria y **escriba la ecuación que especifica el modelo**
- b) Para las trayectorias del proceso AR(2) anterior, grafique la ACF y la PACF
- c) Con los datos correspondientes a la realización del proceso AR(2) en el 1er ítem. **Grafique** diagramas de dispersión de $(z_t$ vs $z_{t-1})$, $(z_t$ vs $z_{t-2})$ y $(z_t$ vs $z_{t-3})$. ¿Qué observa de diferente entre los diagramas de dispersión de este punto y los del punto anterior?.

4.1. a)

Para un modelo AR(2) que sea estacionario con correlación cíclica se puede definir de la siguiente forma:

$$Z_t = \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + a_t$$

Donde:

- $\mu =$ es la media de este proceso donde la tomaremos como $\mu = 0$
- a_t es el ruido blanco, el cual fue el que consideramos anteriormente
- θ_1 y θ_2 son los coeficientes autoregresivos de orden 1 y 2 respectivamente, para este ejercicio los tomaremos $\theta_1 = 0.21$ y $\theta_2 = -0.65$ estos valores fueron escogidos arbitrariamente respetando la condición*
- Z_t es la variable de interés en el tiempo t

*Condición:

$$\begin{aligned} \theta_2 + \theta_1 < 1 \quad y \quad \theta_2 - \theta_1 < 1 \quad y \quad -1 < \theta_2 < 1 \\ -0.65 + 0.21 < 1 \quad y \quad -0.65 - 0.21 < 1 \quad y \quad -1 < -0.65 < 1 \\ -0.44 < 1, \quad y, \quad -0.86 < 1 \quad y \quad -1 < -0.65 < 1 \end{aligned}$$

Ademas de la condición de que le polinomio caracteristico posee al menos 1 raíz compleja para ver el patrón ciclico, esto se verifica de la siguiente forma:

$$(I - \theta_1\beta - \theta_2\beta^2) = (I - 0.21\beta - (-0.65)\beta^2) = (0.65\beta^2 - 0.21\beta + 1)$$

$$\frac{-\theta_1 \pm \sqrt{\theta_1^2 - 4 * I * \theta_2}}{2 * \theta_2} = \frac{0.21 \pm \sqrt{(-0.21)^2 - 4 * 1 * 0.65}}{2 * 0.65} = \frac{0.21 \pm \sqrt{-2.5559}}{1.3}$$

$$\frac{0.21 \pm 1.59872i}{1.3} = \frac{0.21}{1.3} \pm \frac{1.59872i}{1.3} = 0.161538 \pm 1.22978i$$

El polinomio caracteristico posee raices complejas, por lo tanto los valores de nuestra secuencia se espera que se presenten ciclos u osilaciones:

Nuestro modelo AR(2) y polinomio quedaria:

$$Z_t = 0.21Z_{t-1} - 0.65Z_{t-2} + a_t$$

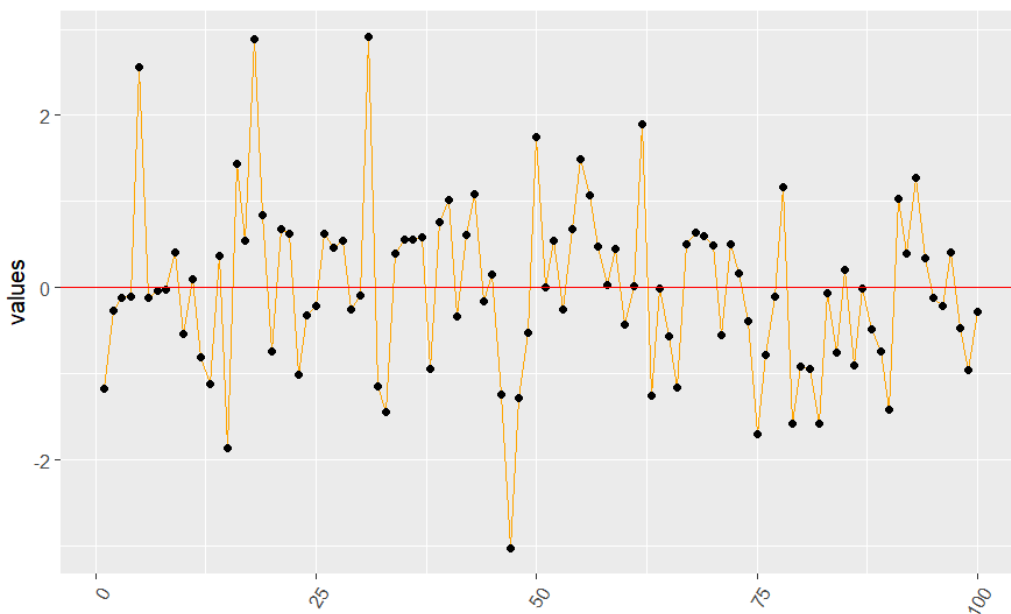
$$(I - 0.21\beta + 0.65\beta^2)$$

Para la realización de nuestro modelo AR(2) debemos escoger valores de nuestra secuencia, los cuales tambien los escogeremos de manera arbitraria.

- $Z_0 = 0.4$.
- $Z_1 = 0.8$.

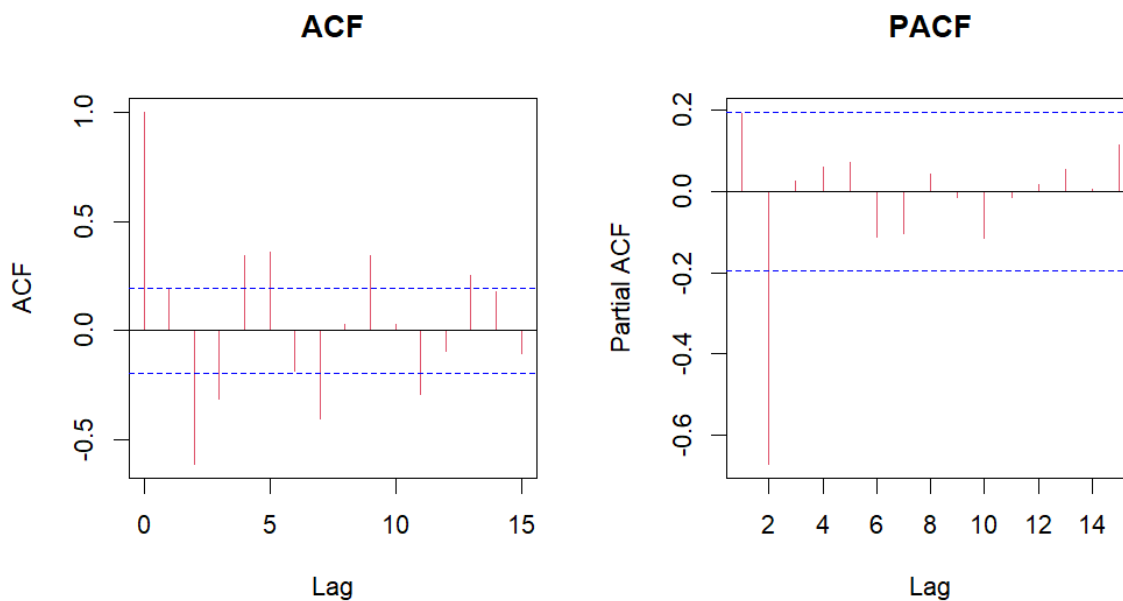
Con todo lo anterior definido, generamos las realizaciones del proceso AR(2), el es un modelo autoregresivo de orden 2, lo que indica que depende linealmente de los valores anteriores en este caso 2, junto al termino del error.

```
theta1 <- 0.21
theta2 <- -0.65
a <- Datos$values
z <- vector()
z[1] <- 0.4
z[2] <- 0.8
for(t in 3:100){
  r <- theta1*z[t-1] + theta2*z[t-2] + a[t]
  z[t] <- r
}
Dat <- data.frame(times,z)
Proceso <- ggplot(Dat, aes(x=times, y=values)) +
  geom_line(color="orange") +
  geom_point() +
  geom_hline(yintercept = 0, color = "red") +
  xlab("") +
  theme(axis.text.x=element_text(angle=60, hjust=1))
Proceso
```



4.2. b)

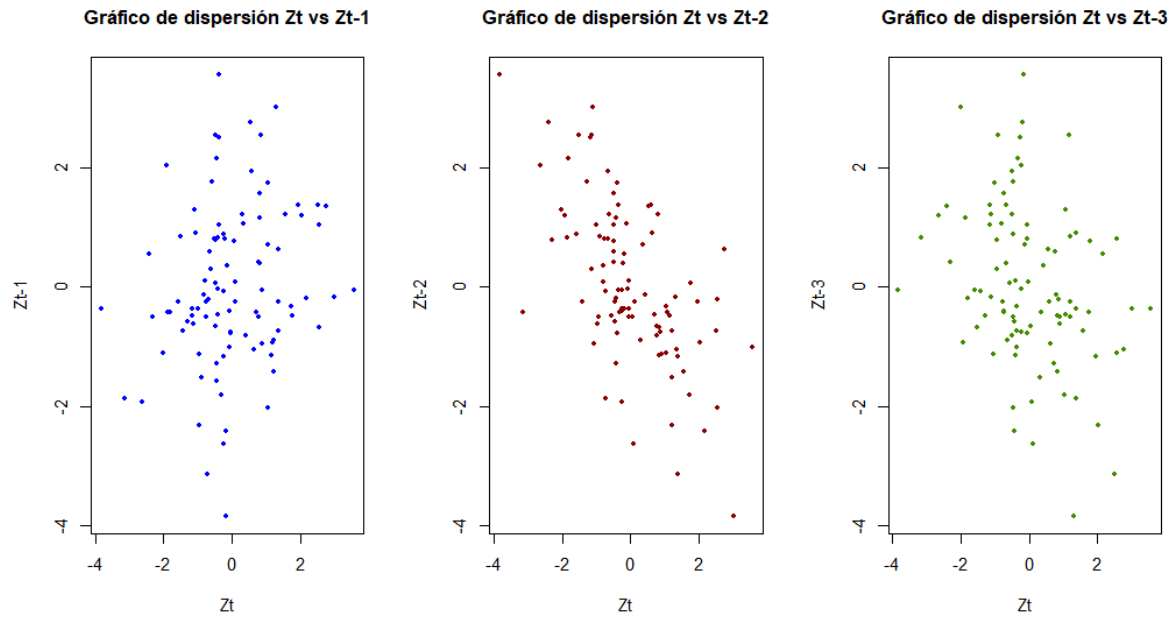
```
par(mfrow=c(1,2))
acf(Dat$z, col = 2, main = "acf", lag.max = 15)
pacf(Dat$z, col = 2, main = "pacf", lag.max = 15)
```



En los correlogramas se puede observar una estructura cíclica de correlación. Además, las autocorrelaciones son significativas, como lo demuestra el hecho de que superan los límites de significancia tanto en la ACF y PACF

4.3. c)

```
par(mfrow=c(1,3))
plot(z[2:100],z[1:99], xlab="Zt", ylab="Zt-1",
     main= "Gráfico de dispersión Zt vs Zt-1", col= "blue",pch = 20)
plot(z[3:100],z[1:98], xlab="Zt", ylab="Zt-2",
     main= "Gráfico de dispersión Zt vs Zt-2", col= "red4",pch = 20)
plot(z[4:100],z[1:97], xlab="Zt", ylab="Zt-3",
     main= "Gráfico de dispersión Zt vs Zt-3", col=
     "chartreuse4",pch = 20)
```



Los gráficos de dispersión revelan una correlación entre los valores de la serie y sus predecesores, mostrando una tendencia lineal positiva entre Z_t vs Z_{t-1} , así como una tendencia lineal negativa entre Z_t vs Z_{t-2} . Este hallazgo contrasta completamente con el modelo MA(1), donde no se observaba ninguna correlación entre las series temporales.