

## Actividad 4

### Estudiante: Maye Mamani Victor Raul

#### Tabla de contenido

CASO 1: VENTAS DE UN RESTAURANTE .....	3
1. IDENTIFICACIÓN .....	3
Análisis de la tendencia y la estacionalidad .....	3
Conclusión: estacionalidad y tendencia .....	5
Análisis de Estacionariedad .....	5
Estacionariedad en Varianza .....	5
Estacionariedad en Media .....	6
Conclusión: Estacionariedad .....	7
Identificación del modelo estacionario .....	7
a) Identificación de las órdenes p y q .....	7
b) Inclusión del termino independiente ( $\delta$ ) o intercepto .....	7
Conclusión: Identificación de los modelos .....	8
2. ESTIMACION .....	8
3. VALIDACIÓN .....	9
3.1. Análisis de los coeficientes estimados .....	9
a. Significación de los coeficientes .....	9
b. Examen de la matriz de correlaciones entre los coeficientes .....	9
c. Condición de Convergencia e Invertibilidad .....	10
d. Análisis de la estabilidad .....	10
3.2. Análisis de los residuos.....	11
a) Media es igual a cero.....	11
b) Homocedasticidad o varianza constante .....	13
c. Ausencia de correlación serial .....	14
d) Contraste de normalidad .....	16
CONCLUSION: ELECCION DEL MEJOR MODELO .....	18
CASO 2: PRODUCCIÓN DE MAIZ.....	18
1.IDENTIFICACION .....	19
Análisis de la tendencia y estacionalidad .....	19
Conclusión estacionalidad y tendencia: .....	19
Análisis de la estacionariedad .....	19
Estacionariedad en varianza .....	19

Estacionariedad en media .....	21
Conclusión de estacionariedad en media y varianza .....	23
Diferenciación.....	23
Conclusión de la diferenciación.....	24
a. Identificación de las órdenes p y q .....	25
b. Inclusión del término independiente o intercepto .....	25
Conclusión de la inclusión del intercepto.....	26
Conclusión: Identificación de los modelos .....	26
2. ESTIMACION .....	26
3. VALIDACION .....	27
3.1. Análisis de los coeficientes estimados .....	27
a. Significación de los coeficientes .....	27
b. Examen de la matriz de correlaciones entre los coeficientes .....	27
c. Condición de Convergencia e Invertibilidad .....	28
d. Análisis de la estabilidad .....	29
3.2 Análisis de residuos .....	30
Homocedasticidad de los residuos.....	33
Correlograma de los residuos.....	34
Prueba de Ljung – Box.....	34
CONCLUSION: .....	37

**LAS INTERPRETACIONES Y CONCLUSIONES DE CADA PASO SE ENCUENTRAN EN ESTE FORMATO**

**Ejemplo:** Según el grafico anterior no hay tendencia

## CASO 1: VENTAS DE UN RESTAURANTE

```
# Librerias necesaria
library(forecast) # Modelo ARIMA
library(tseries) # Para series de tiempo
library(TSA) # Para series de tiempo
library(urca) # Raiz Unitaria
library(ggplot2) # Para hacer gráficos
library(gridExtra)
library(dplyr) # Para la manipulación de datos
library(lmtest) # Inferencia para coeficientes estimados
library(MASS) # Transformacion de Box-Cox
library(nortest) # Pruebas de normalidad
library(strucchange) # Cambio estructural - Test de Chow
library(mFilter)
library(readxl)
library(fitdistrplus)
data <- read_excel("F:\\777--Programacion
repos\\Una\\r\\data\\actividad-04.xlsx", sheet = "datos1")
View(data)
# Gráfica de la serie
data_ts <- ts(data$Ventas, start = 1981, frequency = 1)
plot(data_ts, xlab="Semanas", ylab="Ventas")
```

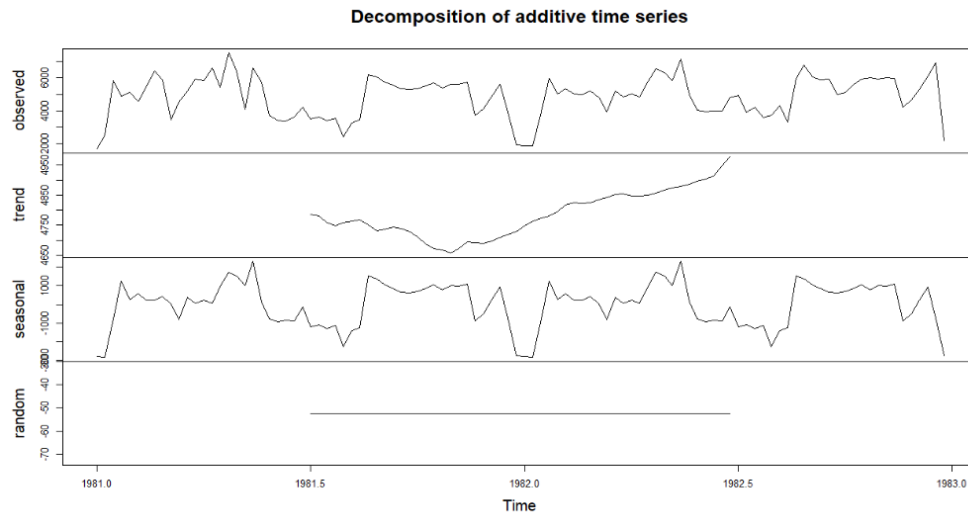


### 1. IDENTIFICACIÓN

#### Análisis de la tendencia y la estacionalidad

```
#Descomposición de la serie
data_des <- decompose(data_ts, type = "additive")
```

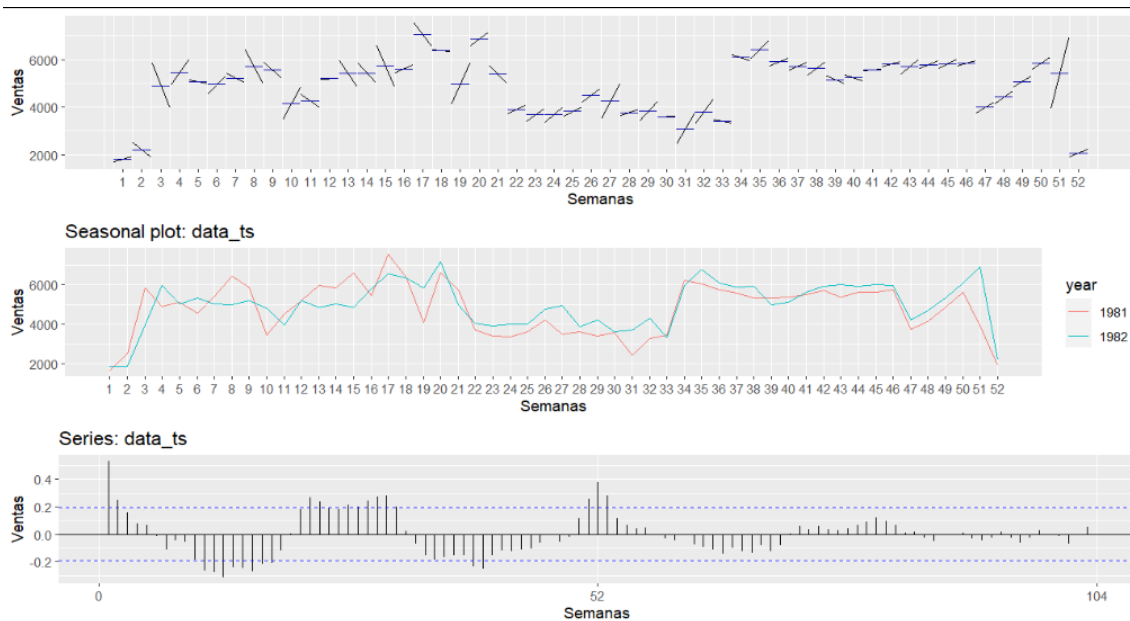
```
plot(data_des, type="l")
```



#Graficos de la serie para identificar estacionalidad

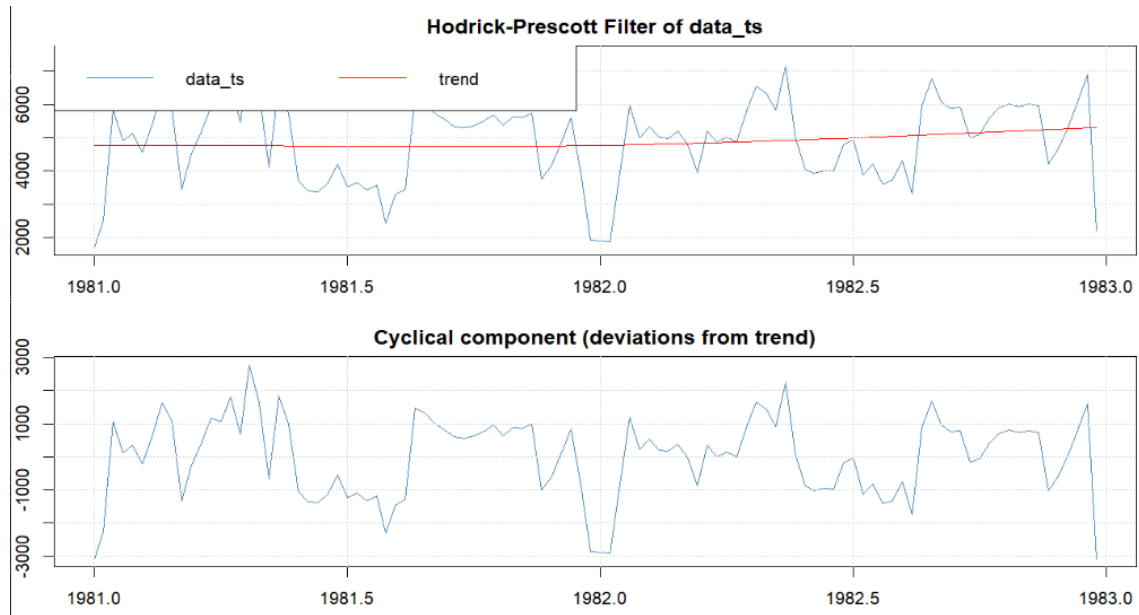
```
plot1 <- ggsubseriesplot(data_ts, xlab = "Semanas", ylab = "Ventas")  
plot2 <- ggseasonplot(data_ts, xlab = "Semanas", ylab = "Ventas")  
plot3 <- ggAcf(data_ts, xlab = "Semanas", ylab = "Ventas")
```

```
grid.arrange(plot1, plot2, plot3, ncol = 1)
```



#Análisis de tendencia

```
lambda_hp <- 129600  
data_hp <- hpfilter(data_ts, type="lambda", freq=lambda_hp)  
plot(data_hp)
```



**Conclusión: estacionalidad y tendencia**

**INTERPRETACION:**

Según los gráficos anteriores NO parece haber una tendencia clara, pero sí hay estacionalidad y parece haber un patrón cíclico que se repite anualmente

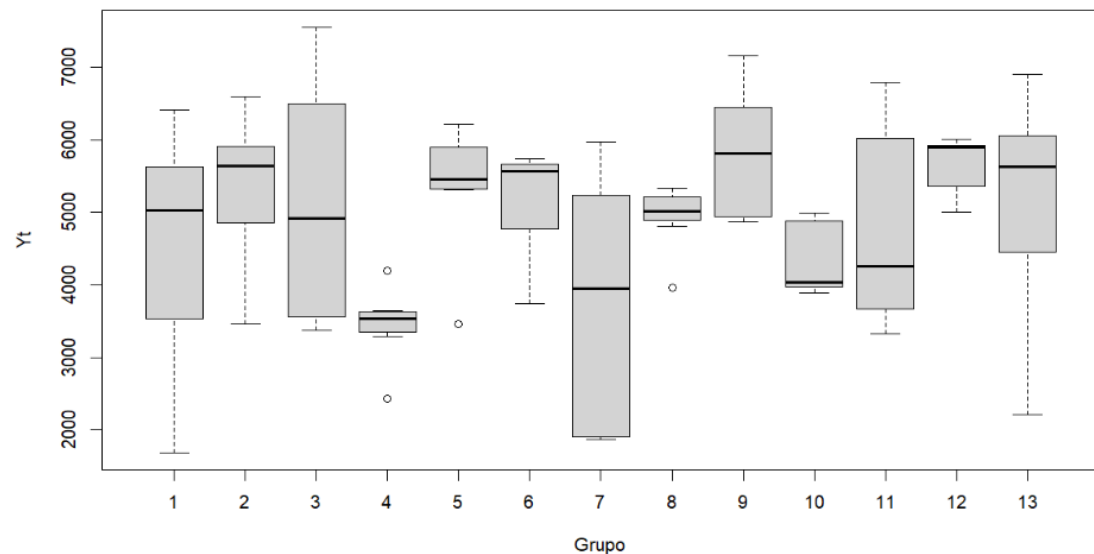
### Análisis de Estacionariedad

#### Estacionariedad en Varianza

#Estacionariedad en varianza

```
Grupo <- rep(1:13, each = 8)
```

```
boxplot(data$Ventas ~ Grupo, xlab = "Grupo", ylab="Yt")
```

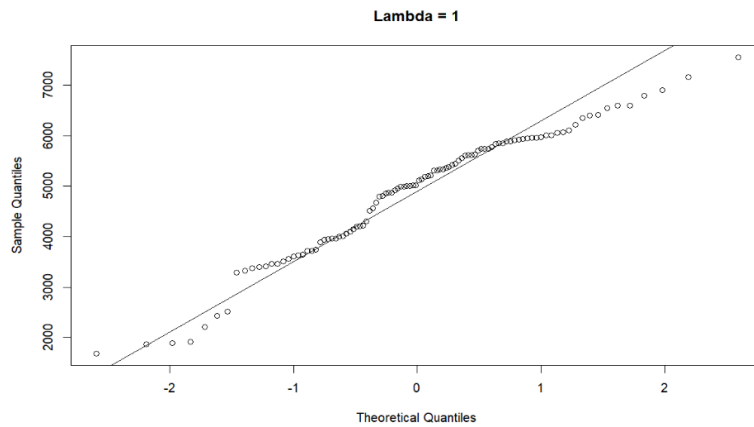


Vemos que la variación no parece ser constante, por lo que probablemente necesitemos transformar la serie.

**Alumno: Maye Mamani Victor Raul**

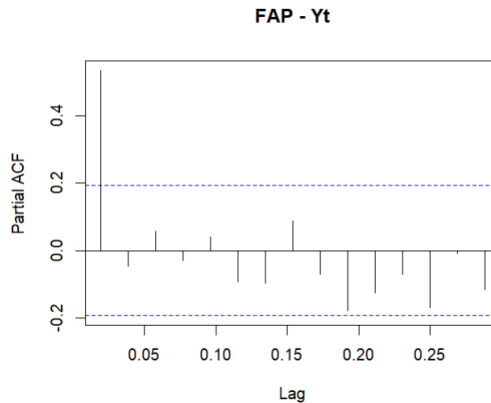
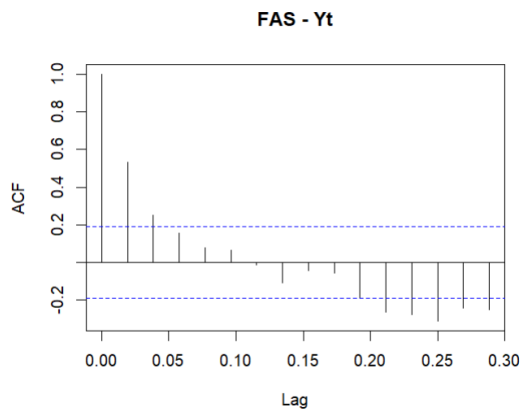
El gráfico muestra la distribución de las ventas de la empresa en el primer trimestre del 2020. La serie de datos se ha transformado para cumplir con los supuestos de normalidad. El resultado es una distribución normal con una media de 1.0 y una desviación estándar de 1.0.

```
qqnorm(data_ts, main = "Lambda = 1") # Yt: Original  
qqline(data_ts)
```



### Estacionariedad en Media

```
FAS <- acf(data_ts, lag.max = 15, main="FAS - Yt", level = 0.95)  
FAP <- pacf(data_ts, lag.max = 15, main="FAP - Yt", level = 0.95)
```



### INTERPRETACION:

Los coeficientes de autocorrelación simple (FAS) decrecen hacia cero rápidamente de forma oscilatoria, además, el primer coeficiente de autocorrelación parcial (FAP) es menor a 0.9; por tanto, esta serie de ventas presenta altos indicios de ser estacionaria.

```
Residual standard error: 1006 on 99 degrees of freedom
Multiple R-squared: 0.2322, Adjusted R-squared: 0.2167
F-statistic: 14.97 on 2 and 99 DF, p-value: 2.093e-06
```

Value of test-statistic is: -5.2981 14.0355

Critical values for test statistics:

	1pct	5pct	10pct
tau2	-3.46	-2.88	-2.57
phi1	6.52	4.63	3.81

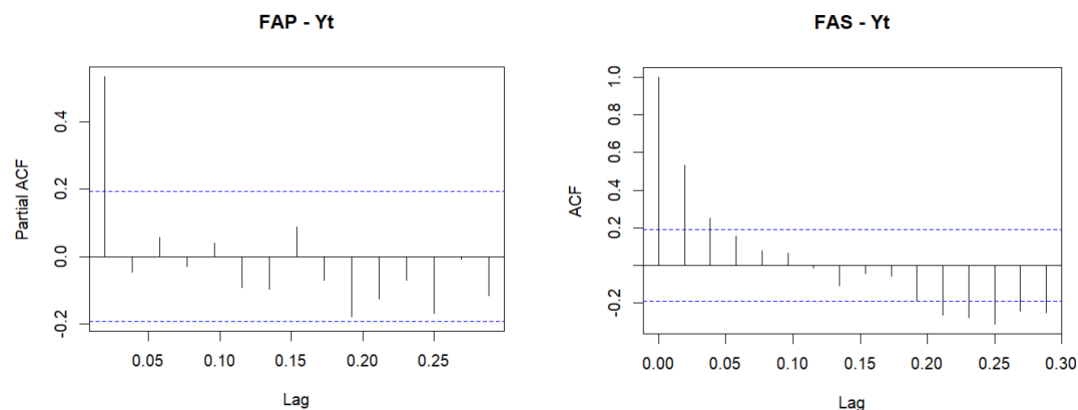
### Conclusión: Estacionariedad

#### INTERPRETACION:

De los resultados, el valor de T calculado = -5.2981, T crítico = -2.8888 por lo tanto, aceptamos la hipótesis nula ( $H_0$ ) de la existencia de raíz unitaria, es decir que la serie es estacionaria en media y varianza

### Identificación del modelo estacionario

#### a) Identificación de las órdenes p y q



La FAS estimada muestra una estructura infinita decreciendo en forma de onda seno-coseno amortiguada que sugiere un modelo ARMA(p, q) con parte autorregresiva de orden = 1.

Los tres primeros coeficientes del FAS son significativamente distintos de cero y el resto no. Según esta interpretación el modelo adecuado sería un AR(1) y MA(3).

#### b) Inclusión del termino independiente ( $\delta$ ) o intercepto

#incluir el intercepto

```
Z <- mean(data_ts)
```

```
Co <- var(data_ts)
```

```
Tn <- length(data_ts)
```

```
Ta <- Tn - 1
```

```
Sigma <- Co/Ta
```

```
t <- Z/Sigma
```

```
tt <- qt(1-0.05/2,Ta-1)
pruebaT <- c(t, tt)
names(pruebaT) <- c("t-calculado", "t-critico")
pruebaT
```

```
> pruebaT
t-calculado t-critico
0.3256219   1.9834953
```

### INTERPRETACION:

Vemos que  $T_{calculado} < T_{critico}$ . Entonces se acepta la hipótesis nula, por lo cual no incluiremos el intercepto

### Conclusión: Identificación de los modelos

Resumiendo, se proponen los siguientes modelos para la serie estacionaria:

$$AR(1) = ARIMA(1, 0, 0) = Z_t = \phi_1 Z_{t-1} + a_t$$

$$MA(2) = ARIMA(0, 0, 2) = Z_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$$

$$ARMA(1, 2) = ARIMA(1, 0, 2) = Z_t = \phi_1 Z_{t-1} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$$

## 2. ESTIMACION

```
mod1 <- Arima(data_ts, order = c(1, 0, 0), include.constant = FALSE)
coeftest(mod1)
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z )
ar1	0.96900	0.01951	49.668	< 2.2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

$$Z_t = 0.96900Z_{t-1} + a_t$$

```
mod2 <- Arima(data_ts, order = c(0, 0, 2), include.constant = FALSE)
coeftest(mod2)
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z )
ma1	1.330084	0.066399	20.032	< 2.2e-16 ***
ma2	0.713893	0.069037	10.341	< 2.2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

$$Z_t = a_t + 1.330084a_{t-1} + 0.713893a_{t-2}$$

```
mod3 <- Arima(data_ts, order = c(1, 0, 2), include.constant = FALSE)
```



coeftest(mod3)

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
ar1	0.9942959	0.0077216	128.7683	< 2e-16	***
ma1	-0.2474759	0.1077573	-2.2966	0.02164	*
ma2	-0.3200939	0.1223653	-2.6159	0.00890	**

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

$$Z_t = 0.9942959Z_{t-1} + a_t - 0.2474759a_{t-1} - 0.3200939a_{t-2}$$

### 3. VALIDACIÓN

#### 3.1. Análisis de los coeficientes estimados

##### a. Significación de los coeficientes

# De las imágenes anteriores analizamos su significancia

Modelo 1: AR(1)

$$\phi_1 = 0.96900 \rightarrow 0.00000 < 0.01, \text{altamente significativo}$$

Modelo 2: MA(2)

$$\theta_1 = 1.330084 \rightarrow 0.00000 < 0.01, \text{altamente significativo}$$

$$\theta_2 = 0.713893 \rightarrow 0.00000 < 0.01, \text{altamente significativo}$$

Modelo 3: ARMA(1,2)

$$\phi_1 = 0.9942959 \rightarrow 0.00000 < 0.01, \text{altamente significativo}$$

$$\theta_1 = -0.2474759 \rightarrow 0.02164 < 0.05, \text{significativo}$$

$$\theta_2 = -0.3200939 \rightarrow 0.1223653 > 0.05, \text{NO es significativo}$$

##### b. Examen de la matriz de correlaciones entre los coeficientes

vcov(mod1)

vcov(mod2)

vcov(mod3)

```
> vcov(mod1)
      ar1
ar1 0.0003806271
> vcov(mod2)
      ma1      ma2
ma1 0.004408807 0.002869561
ma2 0.002869561 0.004766068
> vcov(mod3)
      ar1      ma1      ma2
ar1 5.962296e-05 -0.0003349812 -0.0004350626
ma1 -3.349812e-04 0.0116116320 0.0004372848
ma2 -4.350626e-04 0.0004372848 0.0149732564
> |
```

### INTERPRETACION:

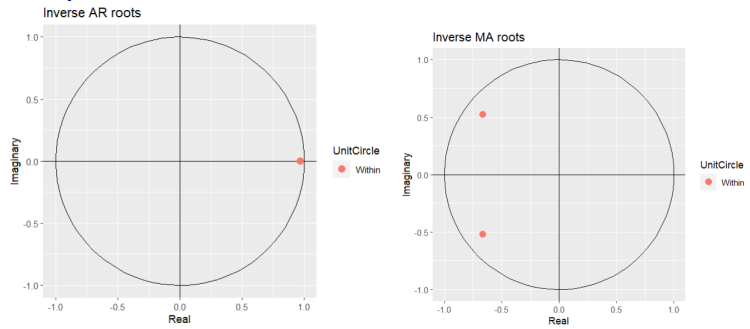
Se observa claramente que ningún coeficiente está próximo ni cercano a 0.9, por tanto, podemos indicar que no hay problema de multicolinealidad en los modelos propuestos.

#### c. Condición de Convergencia e Invertibilidad

```
autoplot(mod1)
```

```
autoplot(mod2)
```

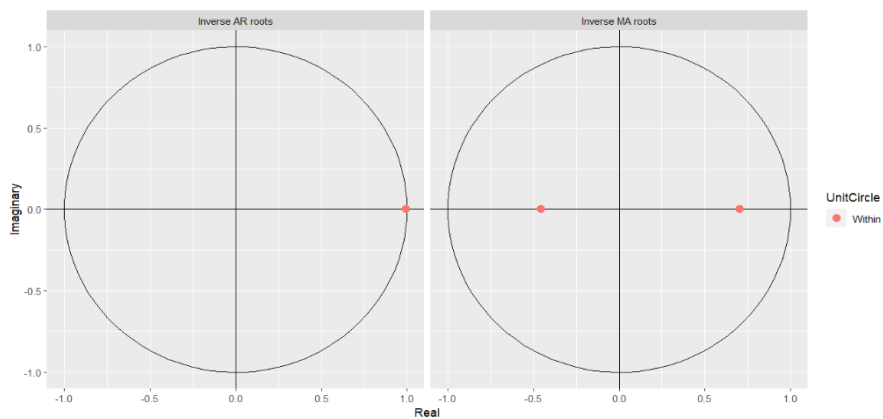
```
autoplot(mod3)
```



### INTERPRETACION:

En la figura de raíces inversas de AR, se observa en el gráfico las raíces características se encuentran dentro del círculo, es decir que cumplen con la condición de estacionariedad para la parte autorregresiva.

En la figura de raíces inversas de MA, se observa en el gráfico las raíces características se encuentran dentro del círculo, es decir que cumplen con la condición de invertibilidad para la parte de media móvil.



### INTERPRETACION:

En la última figura al estar los valores dentro de la circunferencia unitaria es un indicativo de que el modelo se ajusta correctamente. Tanto en su parte AR, como en su parte MA.

#### d. Análisis de la estabilidad

```
Chow_mod1 <- Fstats(mod1$fitted ~ 1, from = 0.43)
```

```
sctest(Chow_mod1)
Chow_mod2 <- Fstats(mod2$fitted ~ 1, from = 0.43)
sctest(Chow_mod2)
Chow_mod3 <- Fstats(mod3$fitted ~ 1, from = 0.43)
sctest(Chow_mod3)
```

```
> Chow_mod1 <- Fstats(mod1$fitted ~ 1, from = 0.49)
> sctest(Chow_mod1)

      supF test

data:  Chow_mod1
sup.F = 3.0989, p-value = 0.1308

> Chow_mod2 <- Fstats(mod2$fitted ~ 1, from = 0.49)
> sctest(Chow_mod2)

      supF test

data:  Chow_mod2
sup.F = 1.4148, p-value = 0.353

> Chow_mod3 <- Fstats(mod3$fitted ~ 1, from = 0.49)
> sctest(Chow_mod3)

      supF test

data:  Chow_mod3
sup.F = 3.7762, p-value = 0.08956
```

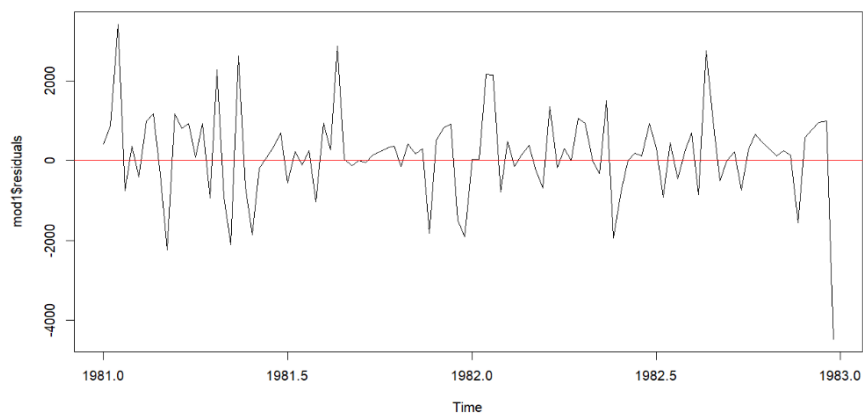
### INTERPRETACION:

En las tres pruebas se acepta la hipótesis nula ( $p > \alpha = 0.05$ ), es decir, existe estabilidad de coeficientes.

### 3.2. Análisis de los residuos

#### a) Media es igual a cero

```
plot(mod1$residuals)
abline(h = 0, col = "red")
```



```
t.test(mod1$residuals, mu = 0)
```

```
One Sample t-test

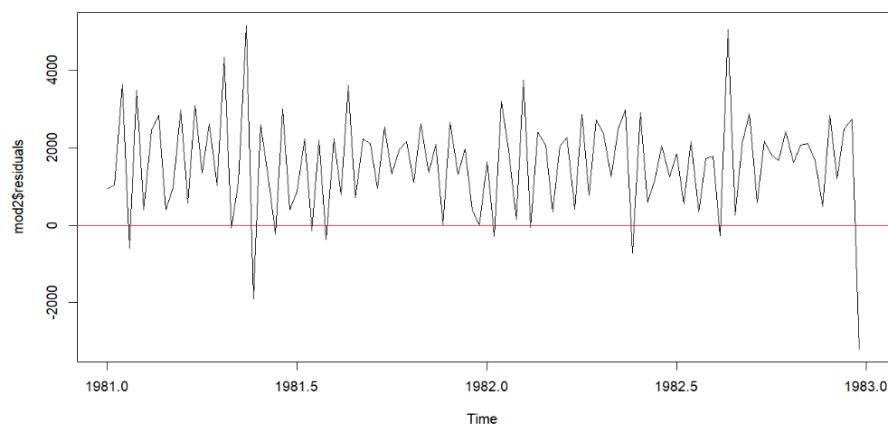
data: mod1$residuals
t = 1.4658, df = 103, p-value = 0.1457
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -56.39116 375.88182
sample estimates:
mean of x
 159.7453
```

#### INTERPRETACION:

Segun el grafico vemos que un buen numero de residuales están en torno a la media igual a cero.

Y confirmando lo anterior  $p=0.1457 > 0.05$ , se acepta  $H_0$ , es decir la media es igual a cero

```
plot(mod2$residuals)
abline(h = 0, col = "red")
```



```
t.test(mod2$residuals, mu = 0)
```

```
One Sample t-test

data: mod2$residuals
t = 12.134, df = 103, p-value < 2.2e-16
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 1327.265 1845.911
sample estimates:
mean of x
 1586.588
```

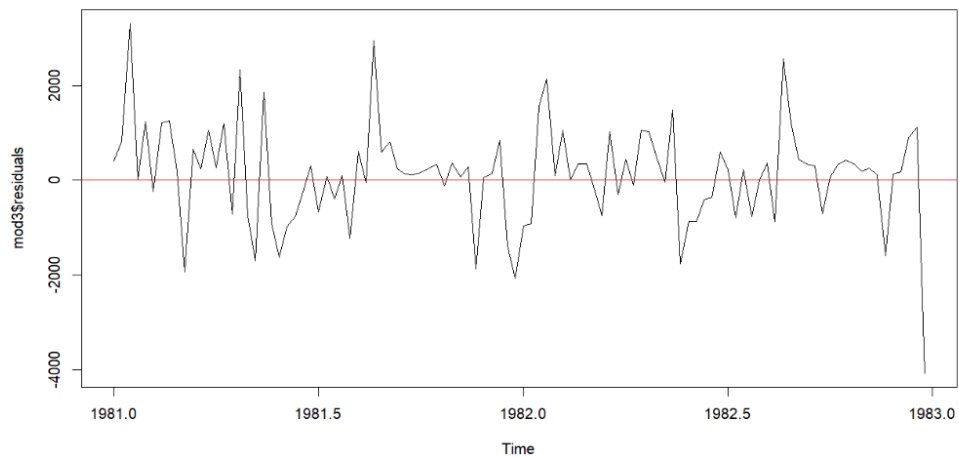
#### INTERPRETACION:

Segun el grafico vemos que los residuales NO están en torno a la media igual a cero.

Y confirmando lo anterior  $p=0.00000 < 0.05$ , se rechaza  $H_0$ , es decir la media no es igual a cero.

```
plot(mod3$residuals)
```

```
abline(h = 0, col = "red")
```



```
t.test(mod3$residuals, mu = 0)
```

```
data: mod3$residuals
t = 1.0655, df = 103, p-value = 0.2892
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -96.2819 319.8295
sample estimates:
mean of x
 111.7738
```

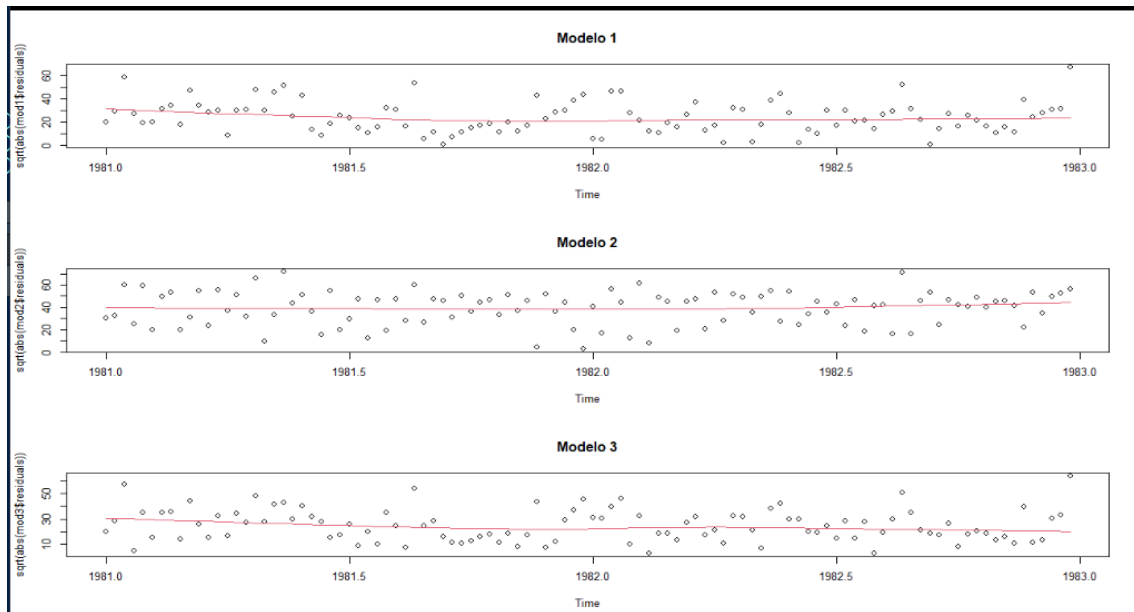
### INTERPRETACION:

Segun el grafico vemos que un buen número de residuales están en torno a la media igual a cero.

Y confirmando lo anterior  $p=0.2892 > 0.05$ , se acepta  $H_0$ , es decir la media es igual a cero

### b) Homocedasticidad o varianza constante

```
par(mfrow = c(3,1))
scatter.smooth(sqrt(abs(mod1$residuals)), lpars=list(col=2), main =
"Modelo 1")
scatter.smooth(sqrt(abs(mod2$residuals)), lpars=list(col=2), main =
"Modelo 2")
scatter.smooth(sqrt(abs(mod3$residuals)), lpars=list(col=2), main =
"Modelo 3")
```



### INTERPRETACION:

Se observa que los datos parecen no presentar una variabilidad considerable, por tanto, será necesario realizar la prueba de Breusch-Pagan para determinar finalmente si las varianzas constantes para los modelos.

```
obs=get(mod1$series)
bptest(resid(mod1)~I(obs-resid(mod1)))
```

```
obs=get(mod2$series)
bptest(resid(mod2)~I(obs-resid(mod2)))
```

```
obs=get(mod3$series)
bptest(resid(mod3)~I(obs-resid(mod3)))
```

studentized Breusch-Pagan test

```
data: resid(mod1) ~ I(obs - resid(mod1))
BP = 0.010868, df = 1, p-value = 0.917
```

studentized Breusch-Pagan test

```
data: resid(mod2) ~ I(obs - resid(mod2))
BP = 0.82445, df = 1, p-value = 0.3639
```

studentized Breusch-Pagan test

```
data: resid(mod3) ~ I(obs - resid(mod3))
BP = 0.22171, df = 1, p-value = 0.6377
```

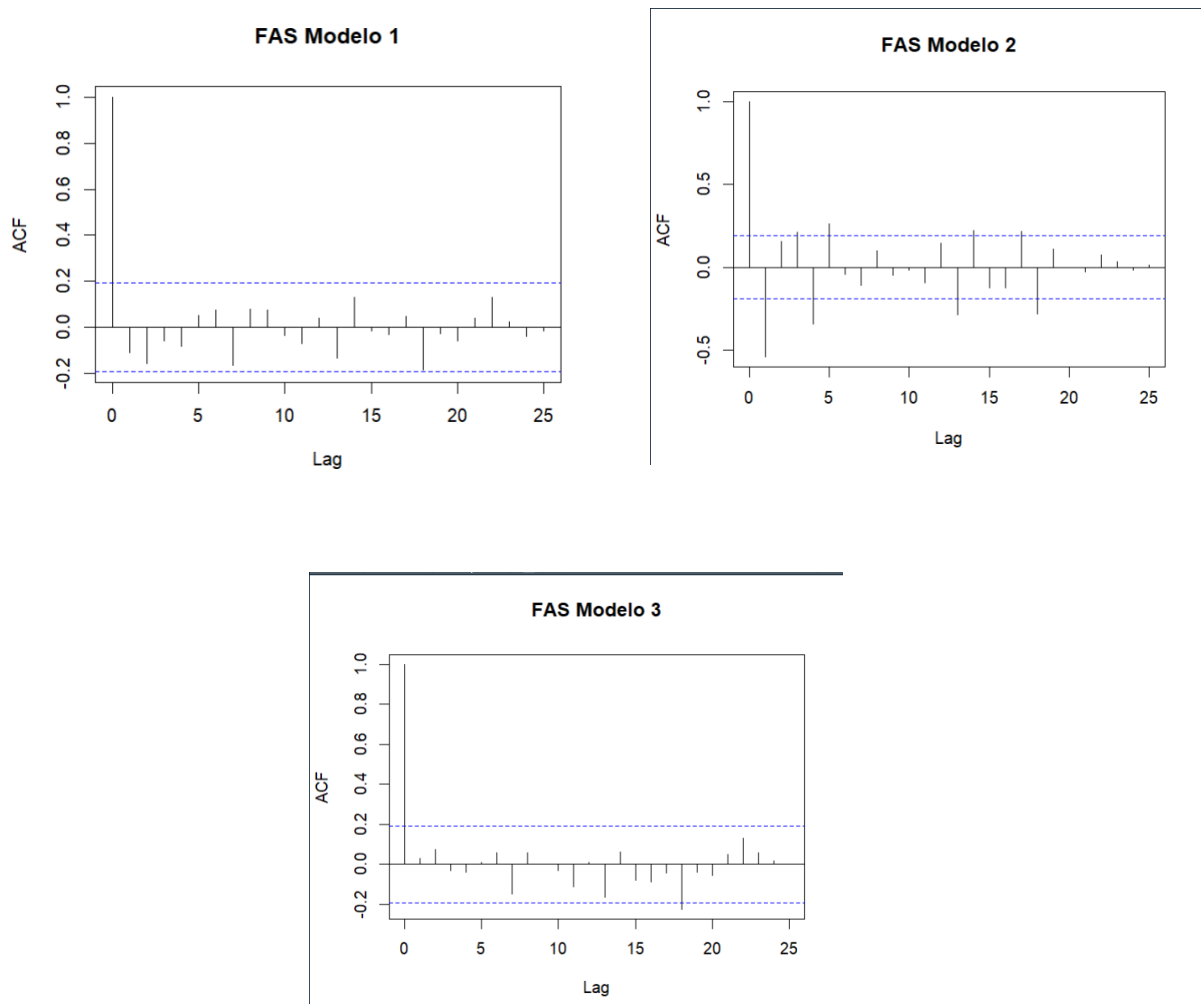
### INTERPRETACION:

El valor de probabilidad (p-valor) asociado al estadístico BP asume un valor de 0.917 para el modelo 1, 0.3639 para el modelo 2 y 0.6377 para el modelo 3. Entonces el p-value de los 3 modelos es mayor a 0.05, con esto podemos decir que los residuales de los 3 modelos son constantes

### c. Ausencia de correlación serial

```
resid_m1 <- as.vector(mod1$residuals)
```

```
resid_m2 <- as.vector(mod2$residuals)
resid_m3 <- as.vector(mod3$residuals)
par(mfrow = c(1,1))
FAS_e.m1 <- acf(resid_m1, lag.max = 25,
                 main="FAS Modelo 1", level = 0.95)
FAS_e.m2 <- acf(resid_m2, lag.max = 25,
                 main="FAS Modelo 2", level = 0.95)
FAS_e.m3 <- acf(resid_m3, lag.max = 25,
                 main="FAS Modelo 3", level = 0.95)
```



### INTERPRETACION:

Se observa que solo un coeficiente del FAS para los modelos 1 y 3 son significativos. Por el contrario el modelo 2 presenta muchos coeficientes que superan los límites de confianza indicando problemas de autocorrelación entre residuales

Prueba de Ljung - Box

```
Box.test(resid_m1,type = "Ljung-Box")
Box-Ljung test
```

data: resid\_m1  
X-squared = 1.3104, df = 1, p-value = 0.2523

```
Box.test(resid_m2,type = "Ljung-Box")  
Box-Ljung test
```

```
data: resid_m2  
X-squared = 31.421, df = 1, p-value = 2.077e-08
```

```
Box.test(resid_m3,type = "Ljung-Box")  
Box-Ljung test
```

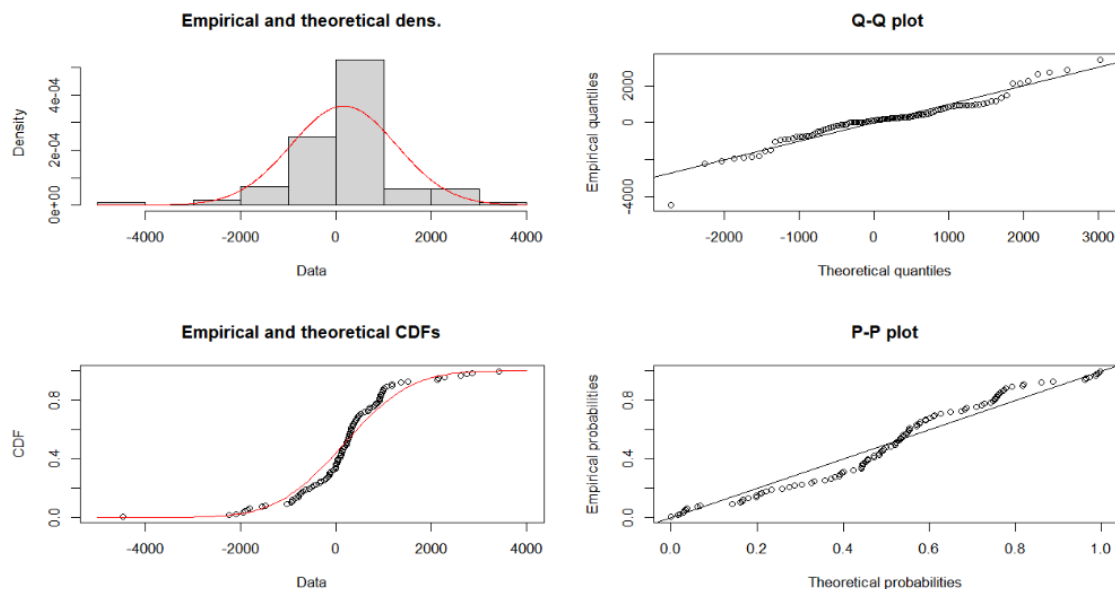
```
data: resid_m3  
X-squared = 0.080731, df = 1, p-value = 0.7763
```

### INTERPRETACION:

Confirmamos lo mencionado en los correlogramas, en los modelos 1 y 3 se acepta la hipótesis nula ( $p\text{-valor} > 0.05$ ) de que los coeficientes de autocorrelación son cero; es decir, los residuos son independientes o están incorrelacionados. Y en el modelo 2 se rechaza la hipótesis nula ( $p\text{-valor} < 0.05$ ) indicándonos que los residuos están correlacionados.

### d) Contraste de normalidad

```
ajuste_m1<-fitdist(data = resid_m1, distr="norm")  
plot(ajuste_m1)  
JB_m1 <- jarque.bera.test(resid_m1)  
JB_m1
```



### Jarque Bera Test

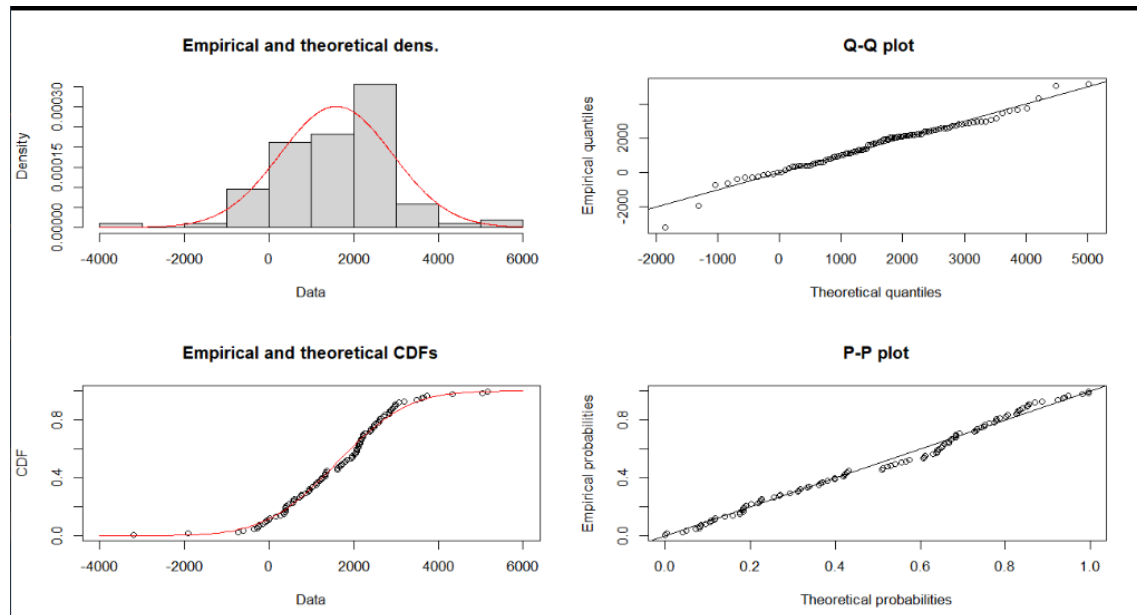
```
data: resid_m1  
X-squared = 40.1, df = 2, p-value = 1.96e-09
```

### INTERPRETACION:



En las figuras se observa que los residuales del modelo 1 presentan indicios de ajuste hacia la distribución normal.

En la prueba JB, como  $p = 0.000000 < 0.05$ , se rechaza  $H_0$ , es decir, los residuos NO se aproximan a una distribución normal.



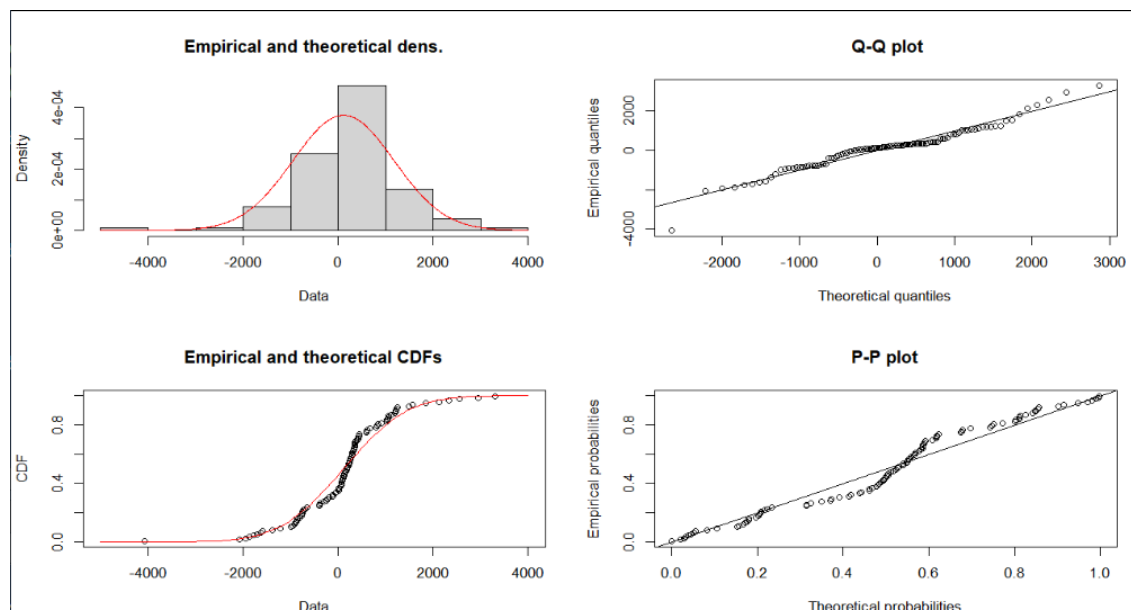
#### Jarque Bera Test

data: resid\_m2  
X-squared = 6.7366, df = 2, p-value = 0.03445

#### INTERPRETACION:

las figuras se observa que los residuales del modelo 1 presentan indicios de ajuste hacia la distribución normal.

En la prueba JB, como  $p = 0.03445 < 0.05$ , se rechaza  $H_0$ , es decir, los residuos NO se aproximan a una distribución normal.



### Jarque Bera Test

```
data: resid_m3  
X-squared = 22.824, df = 2, p-value = 1.106e-05
```

### INTERPRETACION:

En las figuras se observa que los residuales del modelo 1 presentan indicios de ajuste hacia la distribución normal.

En la prueba JB, como  $p = 0.000001 < 0.05$ , se rechaza  $H_0$ , es decir, los residuos NO se aproximan a una distribución normal.

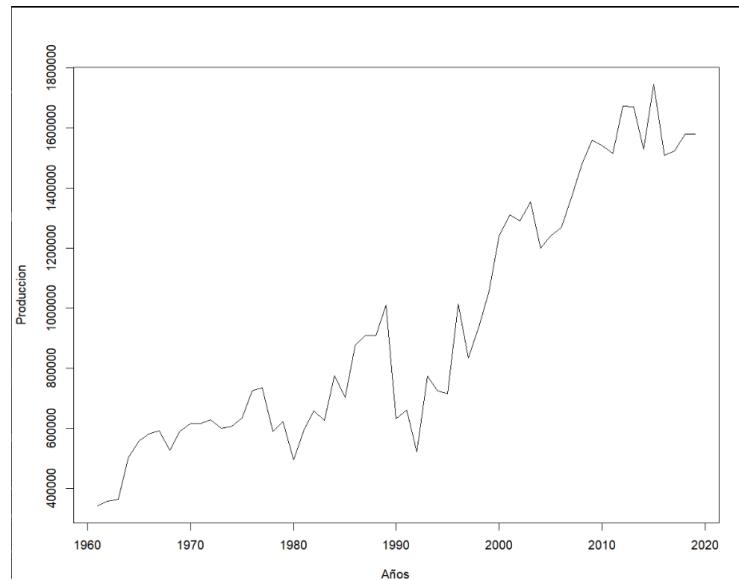
### CONCLUSION: ELECCION DEL MEJOR MODELO

### CONCLUIMOS QUE EL MEJOR MODELO ES EL AR(1) POR LO SIGUIENTE:

- SUS COEFICIENTES SON SIGNIFICATIVOS
- LA MEDIA DE LOS RESIDUALES TIENDE A CERO
- CUMPLE CON LA CONDICIÓN DE ESTACIONALIDAD
- NO TIENE MULTICOLINEALIDAD
- LOS RESIDUALES SON CONSTANTES

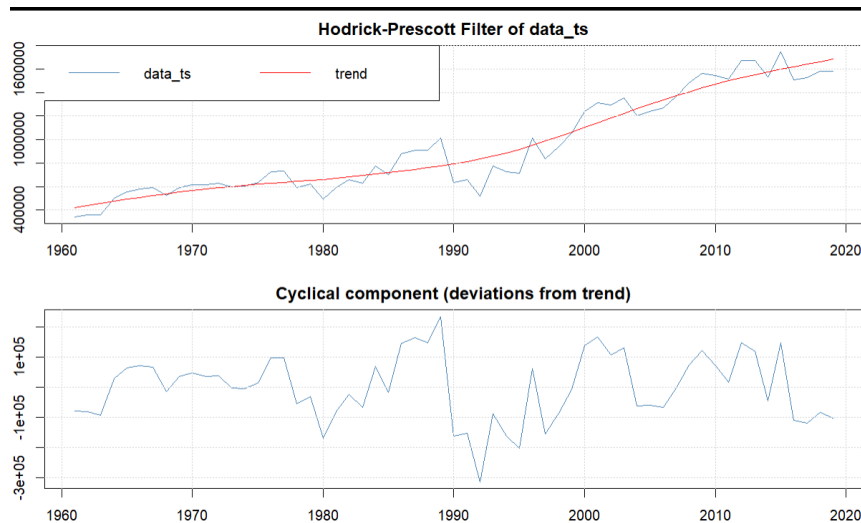
## CASO 2: PRODUCCIÓN DE MAIZ

```
data <- read_excel("F:\\777--Programacion  
repos\\Una\\r\\data\\actividad-04.xlsx", sheet = "datos2")  
View(data)  
# Gráfica de la serie  
data_ts <- ts(data$Produccion, start = c(1961,1), frequency = 1)  
plot(data_ts, xlab="Años", ylab="Produccion")
```



## 1.IDENTIFICACION

### Análisis de la tendencia y estacionalidad



### Conclusión estacionalidad y tendencia:

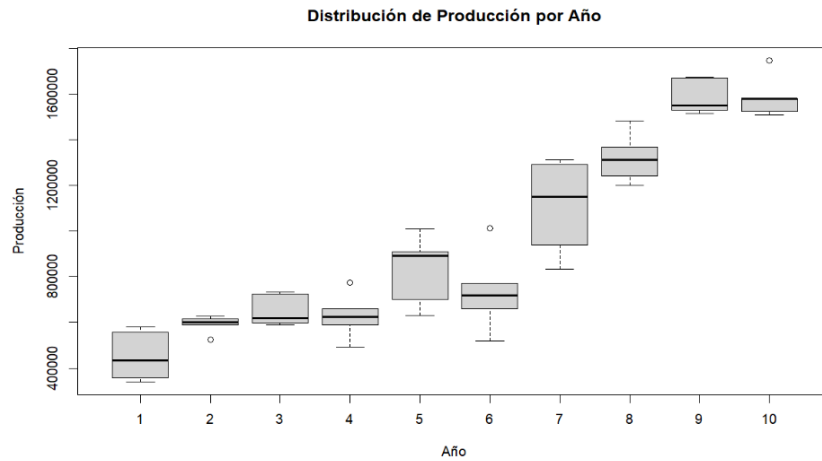
#### INTERPRETACION:

Se observa que la serie si presenta una tendencia creciente, y no tiene estacionalidad dado que la serie es anual y tiene periodicidad menor a un año

### Análisis de la estacionariedad

#### Estacionariedad en varianza

```
Grupo <- rep(1:9, each = 6)
Grupo <- c(Grupo, rep(x = 10,5))
boxplot(data$Produccion ~ Grupo, xlab = "Año", ylab = "Producción",
main = "Distribución de Producción por Año")
```

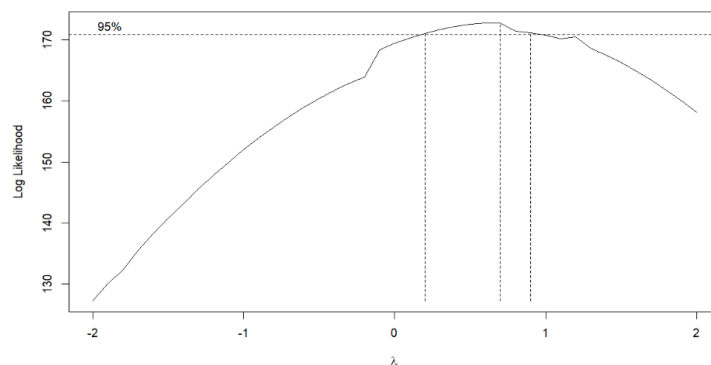


### INTERPRETACION:

Dado el grafico podemos identificar que la serie no parece tener una varianza constante, en otras palabras no aparenta ser estacionaria en varianza.

### VALOR OPTIMO PARA LAMDA

```
b = BoxCox.ar(data_ts)
```



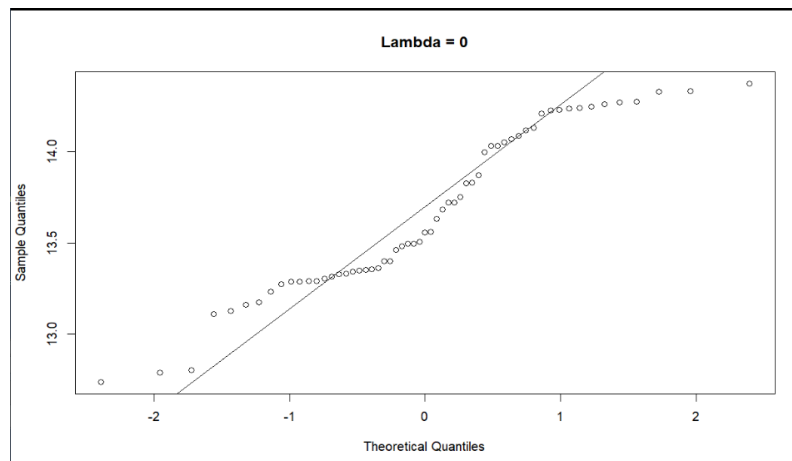
```
lambda <- b$mle
```

```
round(lambda, 0)
```

Observamos que el valor optimo para lamda es 0.7 que se acerca mas a 1, por lo que lambda seria 1. Entonces no hacemos una transformación, ya que sigue siendo la serie original.

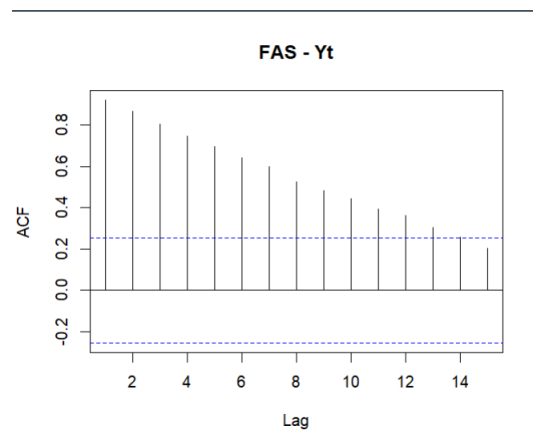
```
qqnorm(data_ts, main = "Lambda = 1") # Yt: Original
```

```
qqline(data_ts)
```

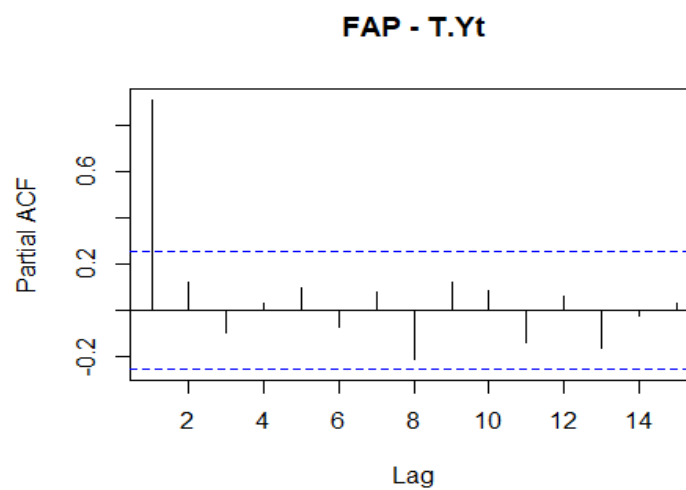


### Estacionariedad en media

```
FAS <- acf(data_ts, lag.max = 15, main="FAS - Yt", level = 0.95)
```



```
FAP <- pacf(data_ts, lag.max = 15, main="FAP - T.Yt", level = 0.95)
```



```
FAP[1]
```

```
##
## Partial autocorrelations of series 'T.Yt', by lag
##
##      1
## 0.921
```

#### INTERPRETACION:

Observamos que el comportamiento de las FAS es decreciente lentamente lo que nos indica que no hay estacionariedad de la serie. Y vemos que el primer coeficiente del FAP es significativo siendo 0.921 mayor a 0.9. Por lo que la serie no es estacionaria.

# Verificación con la prueba de Raíz unitaria de Dickey-Fuller Aumentada.

```
data_adf <- ur.df(data_ts, type="trend", lags = 1)
summary(data_adf)
```

```
#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####
```

Test regression trend

```
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + tt + z.diff.lag)
```

```
Residuals:
    Min       1Q   Median       3Q      Max
-360220  -53502    8013   77557  200586
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.821e+04  3.917e+04   1.997   0.0510 .
z.lag.1      -2.019e-01  9.739e-02  -2.074   0.0430 *
tt           4.604e+03  2.333e+03   1.974   0.0536 .
z.diff.lag   -2.523e-01  1.328e-01  -1.899   0.0630 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 108100 on 53 degrees of freedom
Multiple R-squared:  0.1903,    Adjusted R-squared:  0.1445
F-statistic: 4.153 on 3 and 53 DF,  p-value: 0.01024
```

value of test-statistic is: -2.0737 2.7715 2.1589

Critical values for test statistics:

	1pct	5pct	10pct
tau3	-4.04	-3.45	-3.15
phi2	6.50	4.88	4.16
phi3	8.73	6.49	5.47

#### INTERPRETACION:

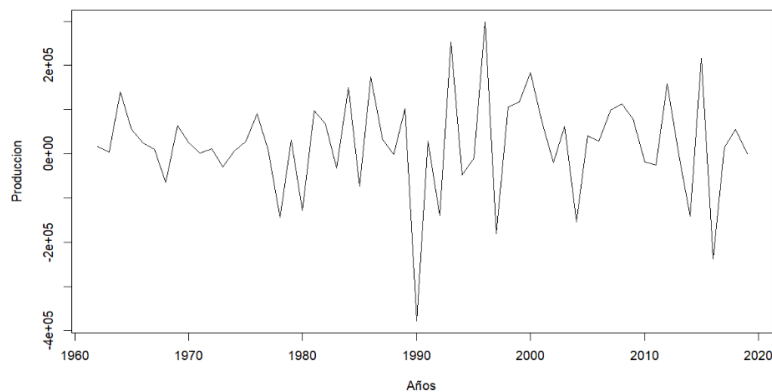
Con un T calculado de -2.0737 > T critico de -3.45. Concluimos que se rechaza la hipótesis nula de la existencia de raíz unitaria, es decir la serie NO es estacionaria en media y varianza

## Conclusión de estacionariedad en media y varianza

Dado lo anterior concluimos que según los métodos graficos y la prueba de raíz unitaria de dickey Fuller que NO HAY ESTACIONARIEDAD EN VARIANZA Y MEDIA

## Diferenciación

```
data_diff <- diff(data_ts)
plot(data_diff, xlab = "Años", ylab="Produccion")
```



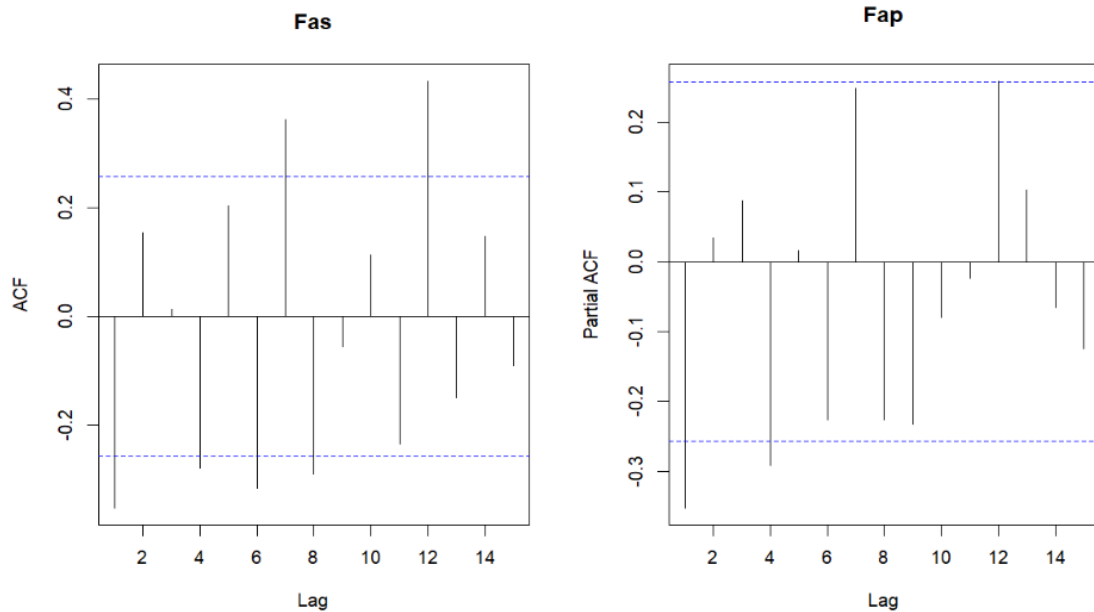
Después de realizar la primera diferencia la serie aparenta ser estacionaria

#Correlograma de la 1ra diferencia

```
par(mfrow =c(1,2))
```

```
FAS <- acf(data_diff, lag.max=15, main="Fas", level = 0.95)
```

```
FAP <- pacf(data_diff, lag.max=15, main="Fap", level =0.95)
```



### INTERPRETACION:

Según el FAS se observa que disminuye rápidamente a 0, y en el FAP vemos que el primer coeficiente es menor a 0.9. Esto nos da a entender que la serie es ESTACIONARIA

#Dickey fuller

```
data_adf <- ur.df(data_diff, type="drift", lags = 1)
summary(data_adf)
```

value of test-statistic is: -5.7842 16.7287

Critical values for test statistics:

	1pct	5pct	10pct
tau2	-3.51	-2.89	-2.58
phi1	6.70	4.71	3.86

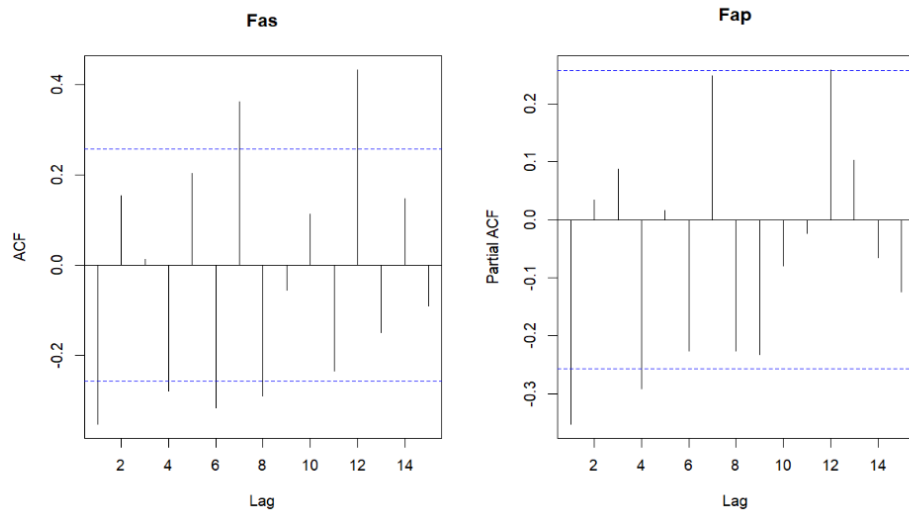
Vemos que T calculado  $-5.7842 < T$  crítico  $-2.89$ . Por lo que haceptamos  $H_0$ , ósea que si hay raíz unitaria. Lo que nos indica que si hay ESTACIONARIEDAD EN MEDIA Y VARIANZA.

### Conclusión de la diferenciación

Dado el análisis anterior usando las autocorrelaciones y la prueba de raíz unitaria se observa que la serie es estacionaria después de aplicar la primera diferencia.



### a. Identificación de las órdenes p y q

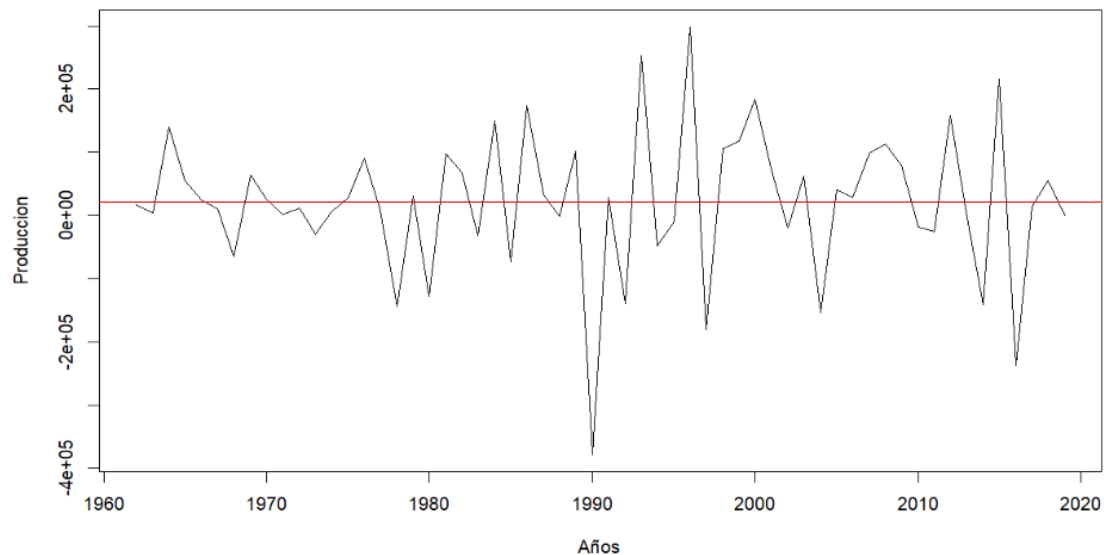


### INTERPRETACION:

Según las FAS y FAP podemos ver que el comportamiento de estas podrían sugerirnos un modelo ARMA(1,2) ya que no muestran un patrón fijo y van seguidos de una mezcla oscilaciones sinusoidales

### b. Inclusión del término independiente o intercepto

```
plot(data_diff, xlab="Años", ylab="Produccion")  
abline(h = mean(data_diff), col = "red")
```



```
Z <- mean(data_diff)  
Co <- var(data_diff)  
Tn <- length(data_ts)  
Ta <- Tn - 1  
Sigma <- Co/Ta
```

```
t <- Z/Sigma
tt <- qt(1-0.05/2,Ta-1)
pruebaT <- c(t, tt)
names(pruebaT) <- c("t-calculado", "t-critico")
pruebaT
```

```
t-calculado    t-critico
9.242922e-05  2.002465e+00
```

### Conclusión de la inclusión del intercepto

El primer grafico nos muestra que la serie estacionaria parece oscilar en torno a una media de 0. Para verificar esto se hizo la prueba estadística donde el  $T_{calculado} < T_{crítico}$ , entonces aceptamos la hipótesis nula. Finalmente concluimos que el intercepto no debe ser incluido en el modelo.

### Conclusión: Identificación de los modelos

Resumiendo, se proponen los siguientes modelos para la serie estacionaria:

$$AR(1) = ARIMA(2, 1, 0) = Z_t = \phi_1 Z_{t-1} + a_t$$

$$MA(1) = ARIMA(0, 1, 2) = Z_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2}$$

$$ARMA(1, 1) = ARIMA(1, 1, 1) =$$

$$Z_t = \phi_1 Z_{(t-1)} + a_t + \theta_1 a_{t-1} + \theta_2 a_{(t-2)}$$

## 2. ESTIMACION

```
#ARIMA(1,1,0)
mod1 <- Arima(data_ts, order = c(1,1,0))
coeftest(mod1)
```

```
z test of coefficients:
      Estimate Std. Error z value Pr(>|z|)
ar1 -0.30209    0.12375  -2.441  0.01464 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$Z_t = -0.30209 Z_{t-1} + a_t$$

Dado la significancia de los coeficientes del modelo AR(1) podemos ver que es significativo para el modelo.

```
#ARIMA(0,1,2)
mod2 <- Arima(data_ts, order = c(0,1,2))
coeftest(mod2)
z test of coefficients:
```

```
      Estimate Std. Error z value Pr(>|z|)
ma1 -0.24076   0.11944 -2.0158  0.04382 *
ma2  0.27947   0.17537  1.5936  0.11102
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$$z_t = -0.24076z_{t-1} + 0.27497z_{t-2} + a_t$$

```
#ARIMA(1,1,2)
mod3 <- Arima(data_ts, order = c(1,1,2))
coeftest(mod3)
```

Dado la significancia de los coeficientes del modelo MA(2) podemos ver que solo MA1 es significativo para el modelo.

```
z test of coefficients:
      Estimate Std. Error z value Pr(>|z|)
ar1 -0.13998   0.32033 -0.4370  0.6621
ma1 -0.12513   0.29434 -0.4251  0.6708
ma2  0.25508   0.18971  1.3446  0.1788
```

$$z_t = -0.13998z_{t-1} + a_t + 0.12513a_{t-1} + 0.25508a_{t-2}$$

**INTERPRETACION:**

Dado la significancia de los coeficientes del modelo ARMA(1,2) podemos ver que ningún coeficiente es significativo para el modelo.

### 3. VALIDACION

#### 3.1. Análisis de los coeficientes estimados

##### a. Significación de los coeficientes

De los coeficientes anteriores analizamos su significancia

Modelo 1: AR(1)

$$\phi_1 = 0.30209 \rightarrow 0.01464 < 0.05, \text{es significativo}$$

Modelo 2: MA(2)

$$\theta_1 = -0.24076 \rightarrow 0.04382 < 0.05, \text{es significativo}$$

$$\theta_2 = -0.27497 \rightarrow 0.11102 > 0.05, \text{NO es significativo}$$

Modelo 3: ARMA(1,2)

$$\phi_1 = 0.13998 \rightarrow 0.6621 > 0.05, \text{NO es significativo}$$

$$\theta_1 = -0.12513 \rightarrow 0.6708 > 0.05, \text{NO es significativo}$$

$$\theta_2 = 0.25508 \rightarrow 0.1788 > 0.05, \text{NO es significativo}$$

##### b. Examen de la matriz de correlaciones entre los coeficientes

```
vcov(mod1)
```

```
ar1  
ar1 0.01531465
```

```
vcov(mod2)
```

```
ma1 ma2  
ma1 0.014264983 0.002097739  
ma2 0.002097739 0.030752962
```

```
vcov(mod3)
```

```
ar1 ma1 ma2  
ar1 0.10261333 -0.08610588 0.02921369  
ma1 -0.08610588 0.08663420 -0.02540416  
ma2 0.02921369 -0.02540416 0.03598994
```

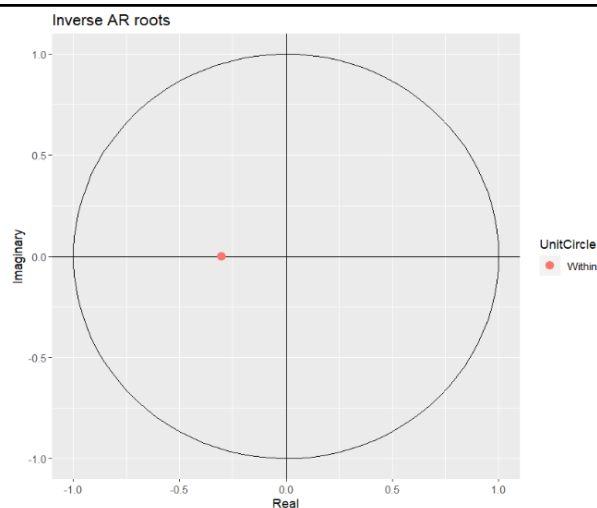
Se observa claramente que ningún coeficiente está próximo ni cercano a 0.9, por tanto, podemos indicar que no hay problema de multicolinealidad en los modelos propuestos.

### c. Condición de Convergencia e Invertibilidad

```
autoplot(mod1)
```

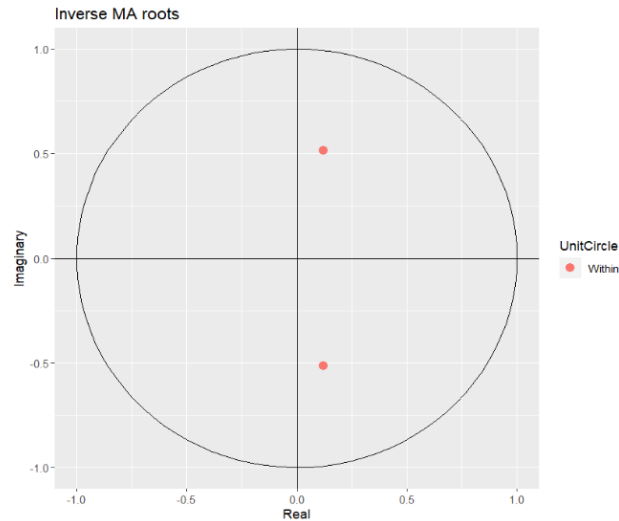
```
autoplot(mod2)
```

```
autoplot(mod3)
```



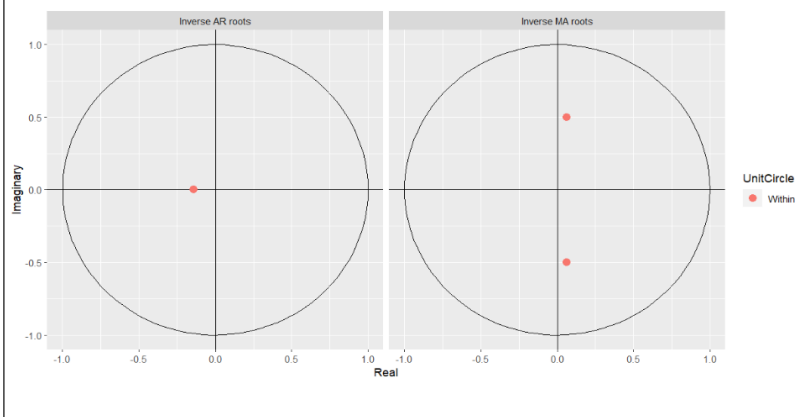
### INTERPRETACION:

En la figura de raíces inversas de AR, se observa en el gráfico las raíces características se encuentran dentro del círculo, es decir que cumplen con la condición de estacionariedad para la parte autorregresiva.



#### INTERPRETACION:

En la figura de raíces inversas de MA, se observa en el gráfico las raíces características se encuentran dentro del círculo, es decir que cumplen con la condición de invertibilidad para la parte de media móvil.



#### INTERPRETACION:

En la última figura al estar los valores dentro de la circunferencia unitaria es un indicativo de que el modelo se ajusta correctamente. Tanto en su parte AR, como en su parte MA.

#### d. Análisis de la estabilidad

```
Chow_mod1 <- Fstats(mod1$fitted ~ 1, from = 0.40)
sctest(Chow_mod1)
```

```
Chow_mod2 <- Fstats(mod2$fitted ~ 1, from = 0.40)
sctest(Chow_mod2)
```

```
Chow_mod3 <- Fstats(mod3$fitted ~ 1, from = 0.40)
sctest(Chow_mod3)
```

supF test

```
data: Chow_mod1
sup.F = 2.2185, p-value = 0.3873

      supF test

data: Chow_mod2
sup.F = 3.8142, p-value = 0.1754

      supF test

data: Chow_mod3
sup.F = 1.7002, p-value = 0.5028
```

En las tres pruebas se acepta la hipótesis nula ( $p > 0.05$ ), es decir, SI existe estabilidad de coeficientes.

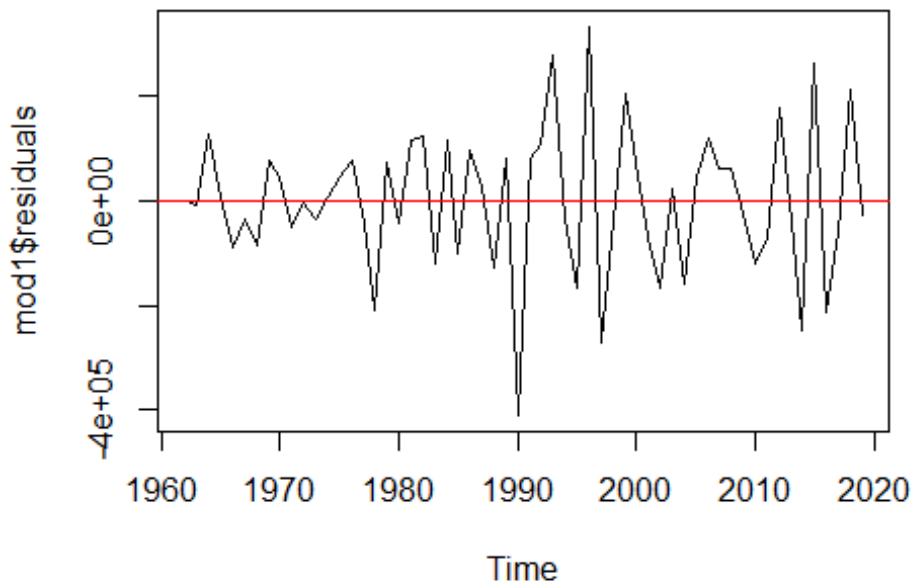
### 3.2 Análisis de residuos

$H_0: E(at) = 0$

$H_i: E(at) \neq 0$

#si el p value es mayor que la significancia aceptamos  $H_0$

```
plot(mod1$residuals)
abline(h = 0, col = "red")
t.test(mod1$residuals, mu = 0)
```



#### One Sample t-test

```
data: mod1$residuals
t = 1.9513, df = 58, p-value = 0.05586
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
```

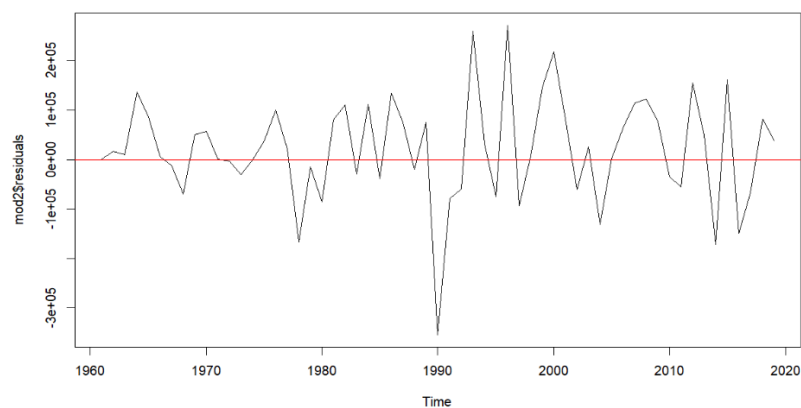
```
-707.3088 55405.9293  
sample estimates:  
mean of x  
27349.31
```

### INTERPRETACION:

Según el grafico vemos que un buen numero de residuales están en torno a la media igual a cero.

Y confirmando lo anterior  $p=0.005586 > 0.05$ , se acepta  $H_0$ , es decir la media es igual a cero

```
plot(mod2$residuals)  
abline(h = 0, col = "red")  
t.test(mod2$residuals, mu = 0)
```



### One Sample t-test

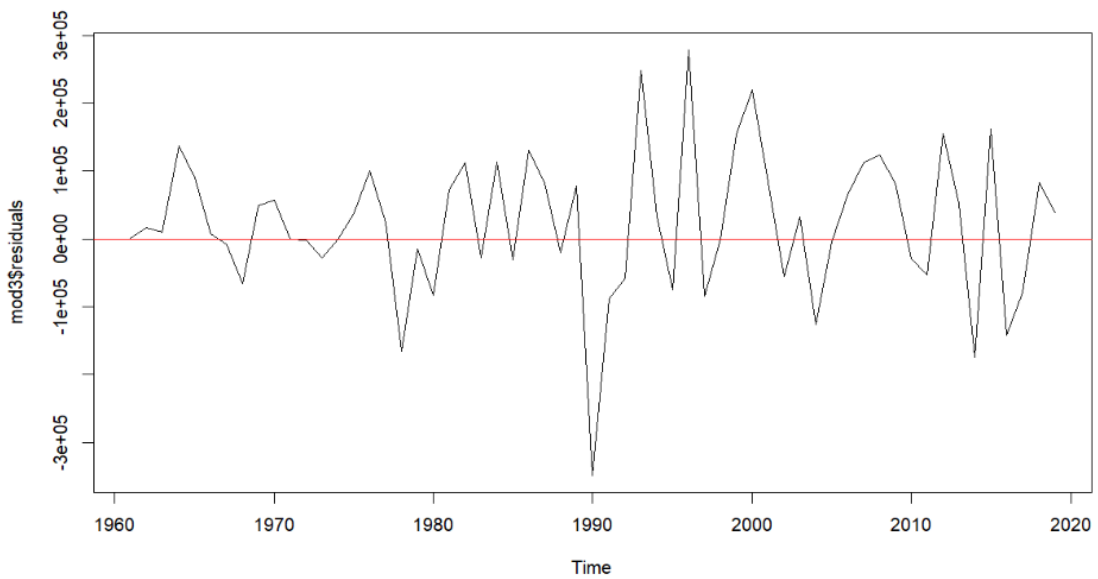
```
data: mod2$residuals  
t = 1.4724, df = 58, p-value = 0.1463  
alternative hypothesis: true mean is not equal to 0  
95 percent confidence interval:  
-7411.469 48648.560  
sample estimates:  
mean of x  
20618.55
```

### INTERPRETACION:

Segun el grafico vemos que un buen numero de residuales están en torno a la media igual a cero.

Y confirmando lo anterior  $p=0.1463 > 0.05$ , se acepta  $H_0$ , es decir la media es igual a cero

```
plot(mod3$residuals)  
abline(h = 0, col = "red")  
t.test(mod3$residuals, mu = 0)
```



```
One Sample t-test
data: mod3$residuals
t = 1.5475, df = 58, p-value = 0.1272
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -6336.359 49516.763
sample estimates:
mean of x
21590.2
```

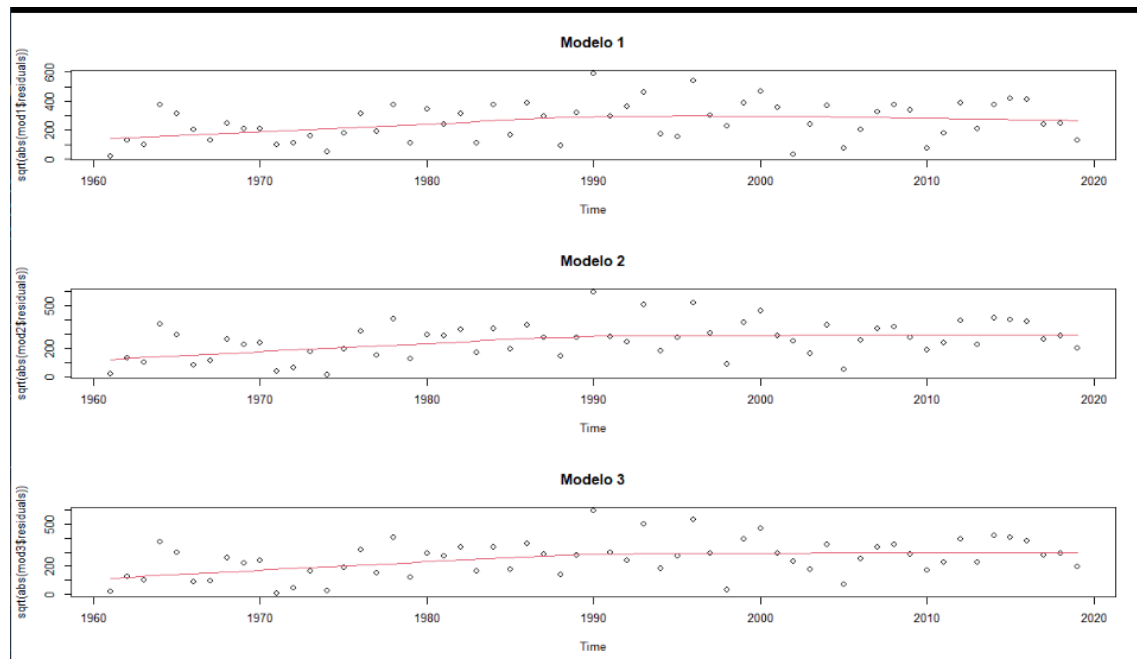
### **INTERPRETACION:**

Segun el grafico vemos que un buen numero de residuales están en torno a la media igual a cero.

Y confirmando lo anterior  $p=0.1272 > 0.05$ , se acepta  $H_0$ , es decir la media es igual a cero



## Homocedasticidad de los residuos



### INTERPRETACION:

Se observa que los datos parecen no presentar una variabilidad considerable, por tanto, será necesario realizar la prueba de Breusch-Pagan para determinar finalmente si las varianzas constantes para los modelos.

### Prueba de Breusch – Pagan

```
obs=get(mod1$series)
bptest(resid(mod1)~I(obs-resid(mod1)))
```

```
studentized Breusch-Pagan test
data: resid(mod1) ~ I(obs - resid(mod1))
BP = 0.57678, df = 1, p-value = 0.4476
```

```
obs=get(mod2$series)
bptest(resid(mod2)~I(obs-resid(mod2)))
```

```
studentized Breusch-Pagan test
data: resid(mod2) ~ I(obs - resid(mod2))
BP = 0.52042, df = 1, p-value = 0.4707
```

```
obs=get(mod3$series)
bptest(resid(mod3)~I(obs-resid(mod3)))
```

```
studentized Breusch-Pagan test
```

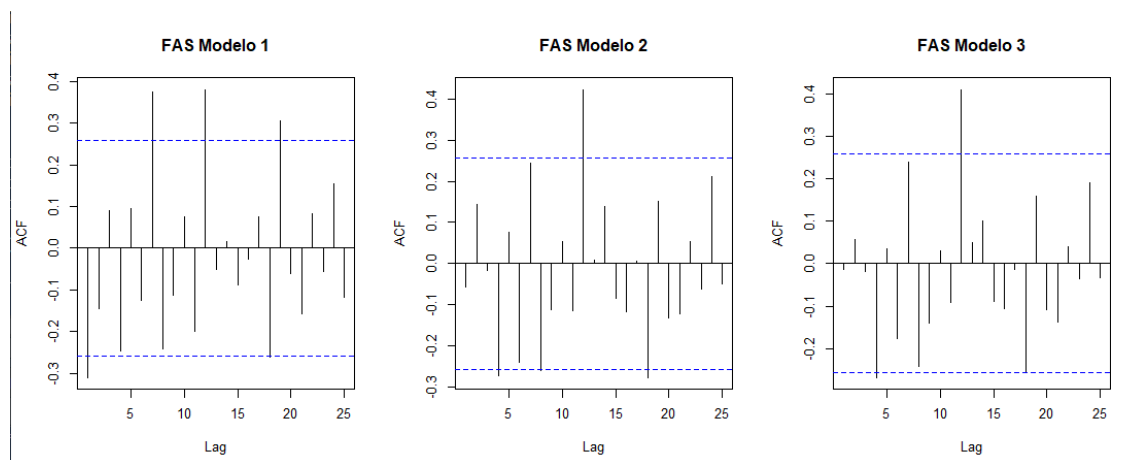
```
data: resid(mod3) ~ I(obs - resid(mod3))  
BP = 0.57112, df = 1, p-value = 0.4498
```

### INTERPRETACION:

El valor de probabilidad (p-valor) asociado al estadístico BP asume un valor de 0.4476 para el modelo 1, 0.4707 para el modelo 2 y 0.4498 para el modelo 3. Entonces el p-value de los 3 modelos es mayor a 0.05, con esto podemos decir que los residuales de los 3 modelos son constantes

### Correlograma de los residuos

```
resid_m1 <- as.vector(mod1$residuals)  
resid_m2 <- as.vector(mod2$residuals)  
resid_m3 <- as.vector(mod3$residuals)  
par(mfrow = c(1,3))  
FAS_e.m1 <- acf(resid_m1, lag.max = 25,  
                 main="FAS Modelo 1", level = 0.95)  
FAS_e.m2 <- acf(resid_m2, lag.max = 25,  
                 main="FAS Modelo 2", level = 0.95)  
FAS_e.m3 <- acf(resid_m3, lag.max = 25,  
                 main="FAS Modelo 3", level = 0.95)
```



### INTERPRETACION:

Se observa que casi la totalidad de los coeficientes del FAS para los modelos 1 y 3 se encuentran dentro de las bandas de no significación, sobre todo los de los primeros retardos. Pero en el modelo 1 se ve varios coeficientes significativos indicando problemas de autocorrelación.

### Prueba de Ljung – Box

```
Box.test(resid_m1,type = "Ljung-Box")  
Box.test(resid_m2,type = "Ljung-Box")
```

```
Box.test(resid_m3,type = "Ljung-Box")
```

```
Box-Ljung test
data: resid_m1
X-squared = 5.8325, df = 1, p-value = 0.01573

Box-Ljung test
data: resid_m2
X-squared = 0.20182, df = 1, p-value = 0.6533

Box-Ljung test
data: resid_m3
X-squared = 0.012917, df = 1, p-value = 0.9095
```

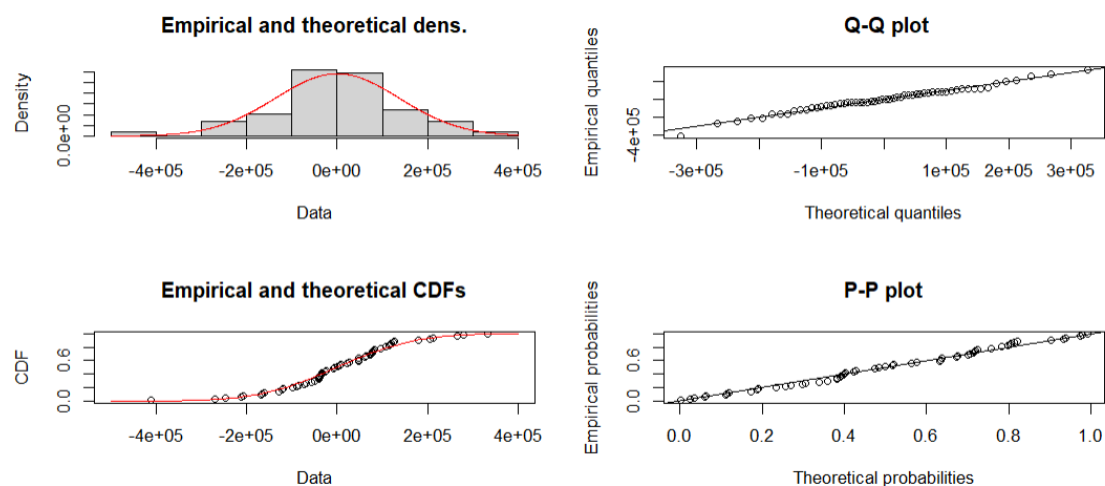
### INTERPRETACION:

Confirmamos lo mencionado en los correlogramas, en los modelos 1 y 3 se acepta la hipótesis nula ( $p\text{-valor} > 0.05$ ) de que los coeficientes de autocorrelación son cero; es decir, los residuos son independientes o están incorrelacionados.

Y en el modelo 1 se rechaza la hipótesis nula indicando que los residuos están correlacionados

### d. Contraste de normalidad

```
ajuste_m1<-fitdist(data = resid_m1, distr="norm")
plot(ajuste_m1)
JB_m1 <- jarque.bera.test(resid_m1)
JB_m1
```

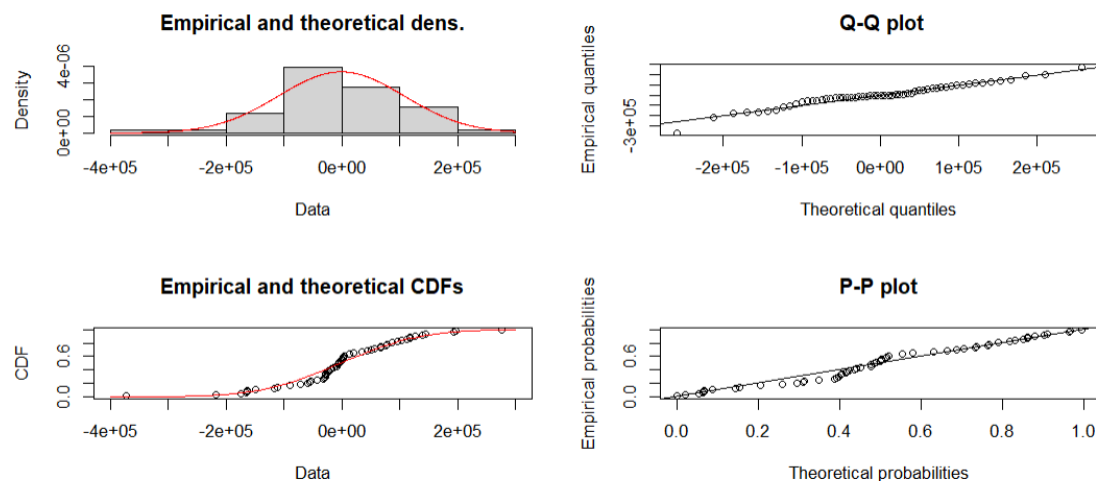


```
Jarque Bera Test
data: resid_m1
X-squared = 1.2796, df = 2, p-value = 0.5274
```

En las figuras se observa que los residuales del modelo 1 presentan indicios de ajuste hacia la distribución normal.

En la prueba JB, como  $p = 0.5274 > 0.05$ , se acepta  $H_0$ , es decir, los residuos se aproximan a una distribución normal.

```
ajuste_m2<-fitdist(data = resid_m2, distr="norm")
plot(ajuste_m2)
JB_m2 <- jarque.bera.test(resid_m2)
JB_m2
```



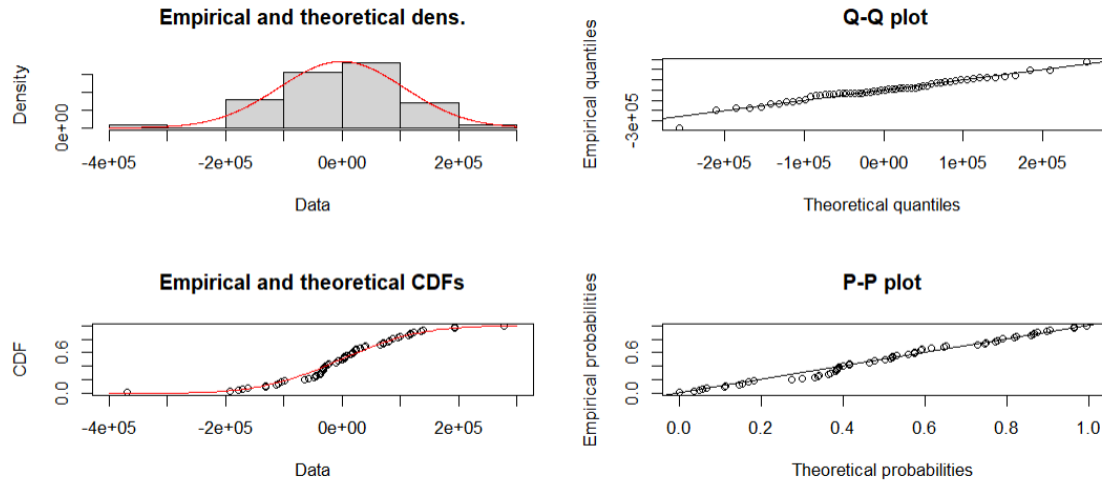
#### Jarque Bera Test

```
data: resid_m2
X-squared = 7.3252, df = 2, p-value = 0.02567
```

En las figuras se observa que los residuales del modelo 1 presentan indicios de ajuste hacia la distribución normal.

En la prueba JB, como  $p = 0.0257 < 0.05$ , se rechaza  $H_0$ , es decir, los residuos NO se aproximan a una distribución normal.

```
ajuste_m3<-fitdist(data = resid_m3, distr="norm")
plot(ajuste_m3)
JB_m3 <- jarque.bera.test(resid_m3)
JB_m3
```



**En las figuras se observa que los residuales del modelo 1 presentan indicios de ajuste hacia la distribución normal.**

**En la prueba JB, como  $p = 0.4199 < 0.05$ , se rechaza  $H_0$ , es decir, los residuos NO se aproximan a una distribución normal.**

### **CONCLUSION:**

**El modelo con mejores métricas fue el fue el AR(1) por las siguientes razones:**

- Los residuos se aproximan a una distribución normal
- Los residuos son independientes
- Cumple con la condición de estacionalidad
- No tiene multicolinealidad
- Los coeficientes son estables