

Índice

1. Chapter 2, Section 2.10: Exercises 3, 8, 10	2
1.1. Exercise 3	2
1.1.1. Solución	2
1.2. Exercise 8	6
1.2.1. Solución	7
1.3. Exercise 10	7
1.3.1. Solución	7
2. Chapter 3, Section 3.7: Exercises 3, 5, 8	9
2.1. Exercise 3	9
2.1.1. Solución	9
2.2. Exercise 5	9
2.3. Exercise 8	11
2.3.1. Solución	11
3. Chapter 6, Section 6.9: Exercise 2.	12
4. Chapter 8, Section 8.11: Exercises 2, 5, 6	12
5. Exercise 3.5	12
5.1. Display and interpret the time series plot for these data	12
5.1.1. Solución	12
5.2. Use least squares to fit a linear time trend to this time series. Interpret the regression output. Save the standardized residuals from the fit for further analysis.	12
5.2.1. Solución	12
5.3. Construct and interpret the time series plot of the standardized residuals from part (b).	13
5.3.1. Solución	13
5.4. Use least squares to fit a quadratic time trend to the wages time series. Interpret the regression output. Save the standardized residuals from the fit for further analysis.	14
5.4.1. Solución	14
5.5. Construct and interpret the time series plot of the standardized residuals from part (d)	15
6. Exercise 3.11	15
6.1. Consider the residuals from a least squares fit of a quadratic time trend . . .	15
6.1.1. Solución	15
6.2. Perform a runs test on the standardized residuals and interpret the results .	16
6.2.1. Solución	16

6.3. Calculate and interpret the sample autocorrelations for the standardized residuals	16
6.3.1. Solución	16
6.4. Investigate the normality of the standardized residuals. Consider histograms and normal probability plots. Interpret the plots.	17
6.4.1. Solución	17

1. Chapter 2, Section 2.10: Exercises 3, 8, 10

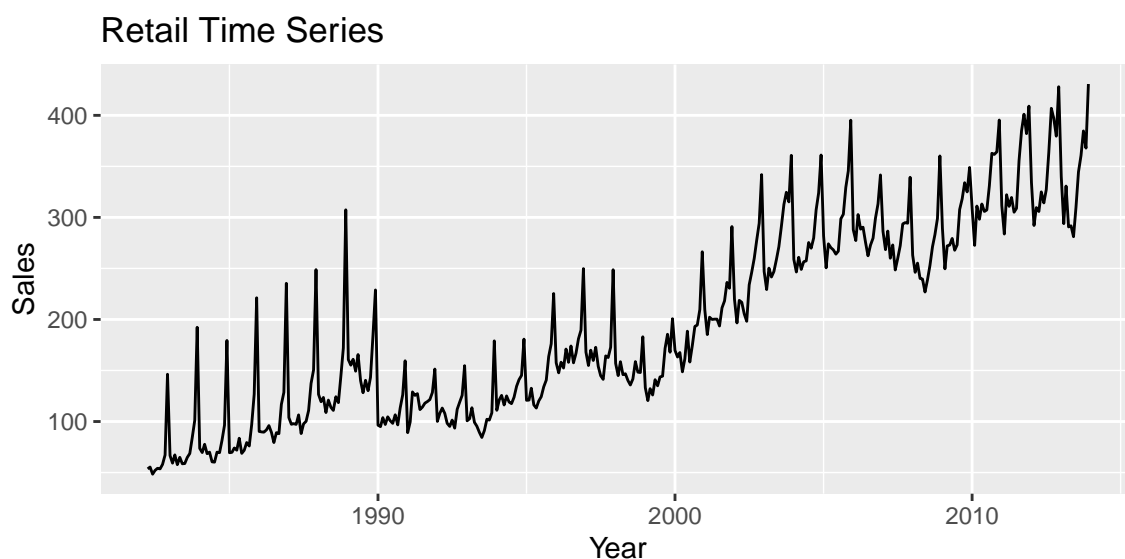
1.1. Exercise 3

Download some monthly Australian retail data from the book website. These represent retail sales in various categories for different Australian states, and are stored in a MS-Excel file.

1.1.1. Solución

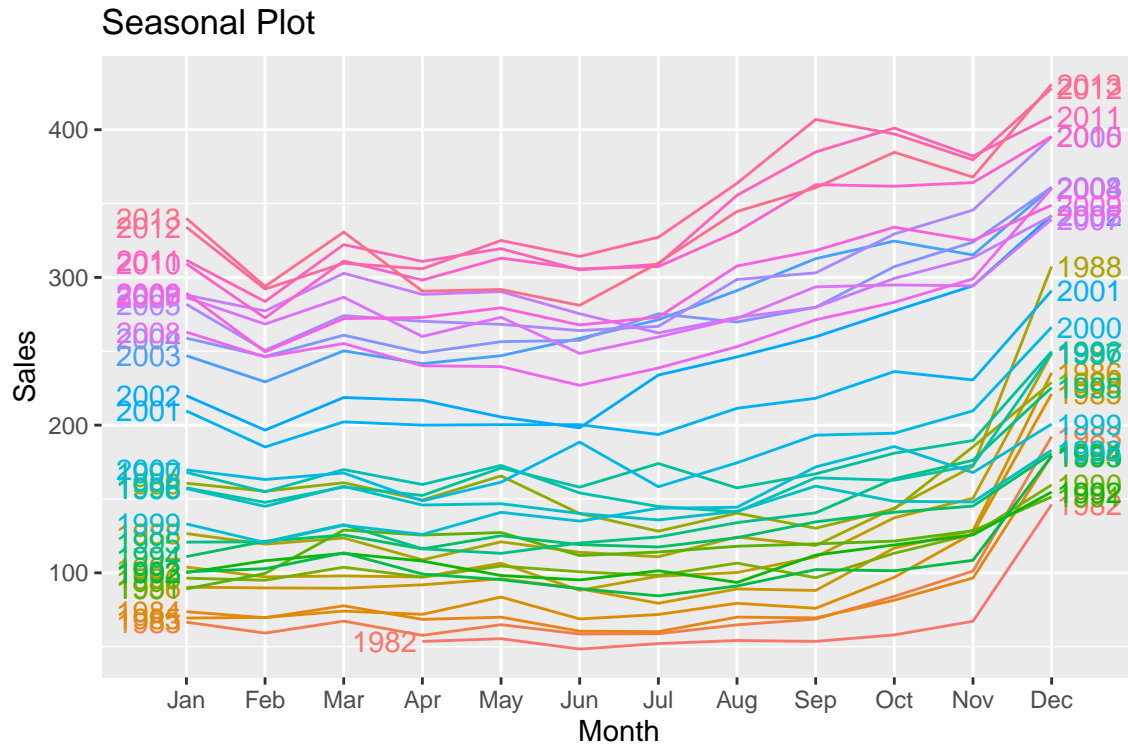
```
retaildata <- readxl::read_excel("retail.xlsx", skip=1)
myts <- ts(retaildata[, "A3349337W"],
           frequency=12, start=c(1982,4))

autoplot(myts) +
  ggtitle("Retail Time Series") +
  xlab("Year") + ylab("Sales")
```



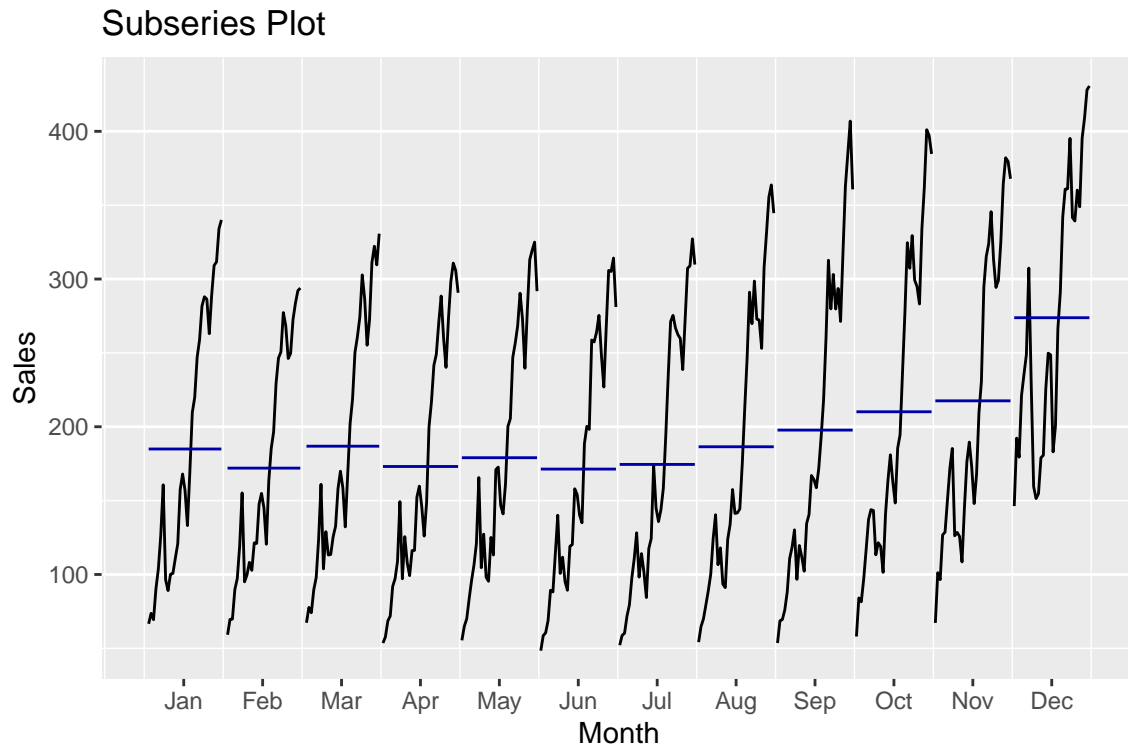
En este gráfico del *retail* se observa una estacionalidad cada 10 años en los datos que va ascendiendo.

```
ggseasonplot(myts, year.labels=TRUE, year.labels.left=TRUE) +
  ggtitle("Seasonal Plot") +
  ylab("Sales")
```



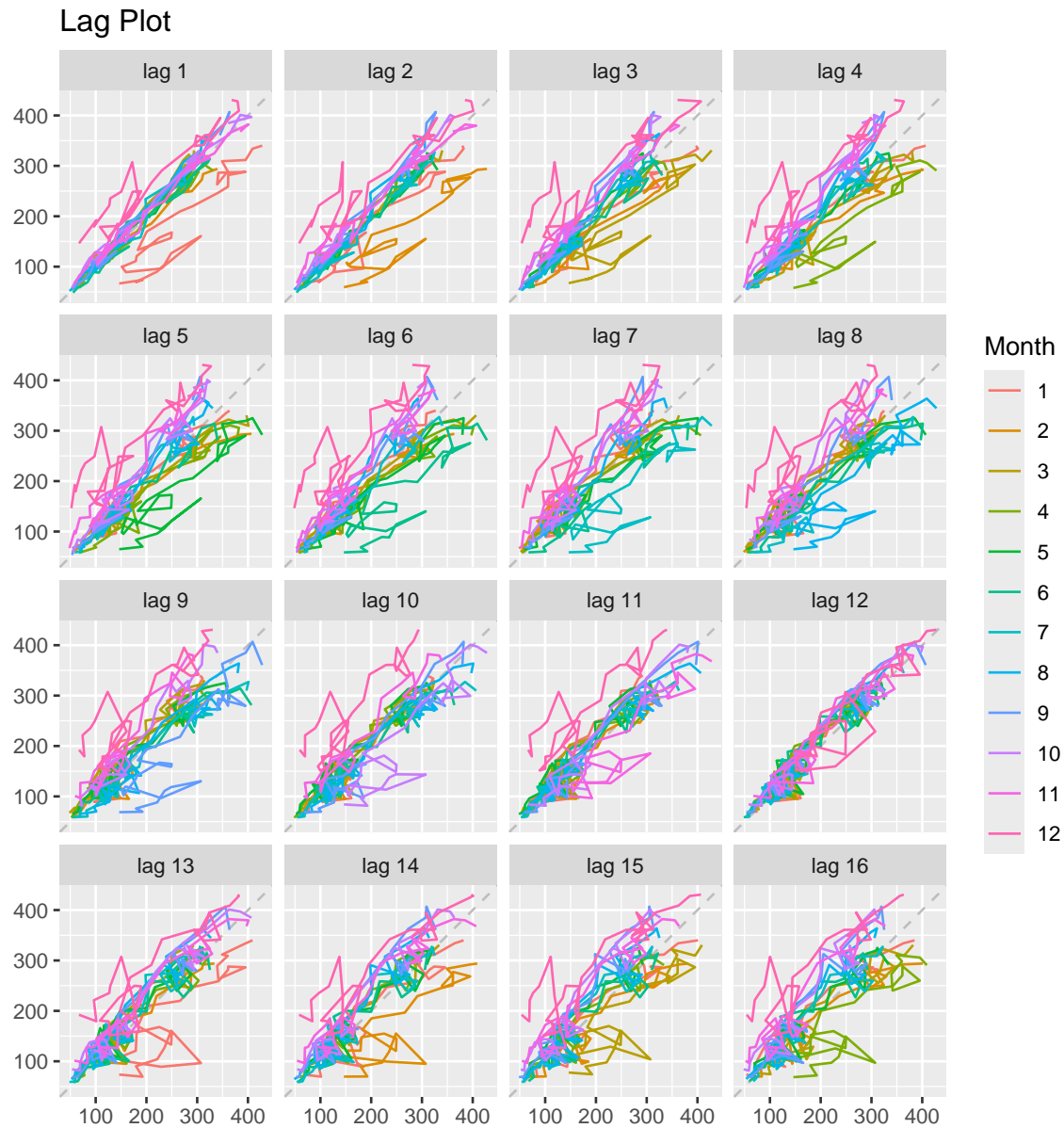
De acuerdo al gráfico estacional, se puede ver que todos los años tienen una tendencia positiva.

```
ggsubseriesplot(myts) +
  ggtitle("Subseries Plot") +
  ylab("Sales")
```



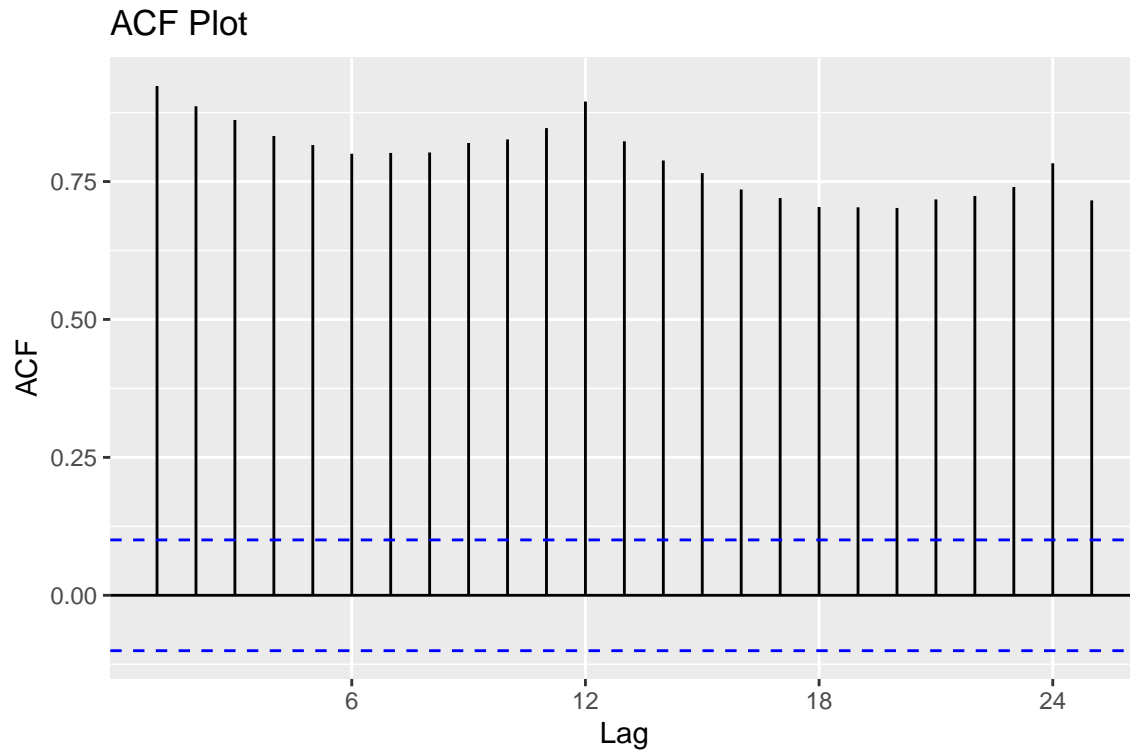
Las ventas durante todos los meses de todos los años tienen un promedio similar, sin embargo, en diciembre estas se disparan, puede deberse a que sea temporada de regalos y festividades.

```
gglagplot(myts) +  
  ggtitle("Lag Plot")
```



Para el *lag plot* podemos ver que en el 12 y el 1 los datos presentan una autocorrelación alta.

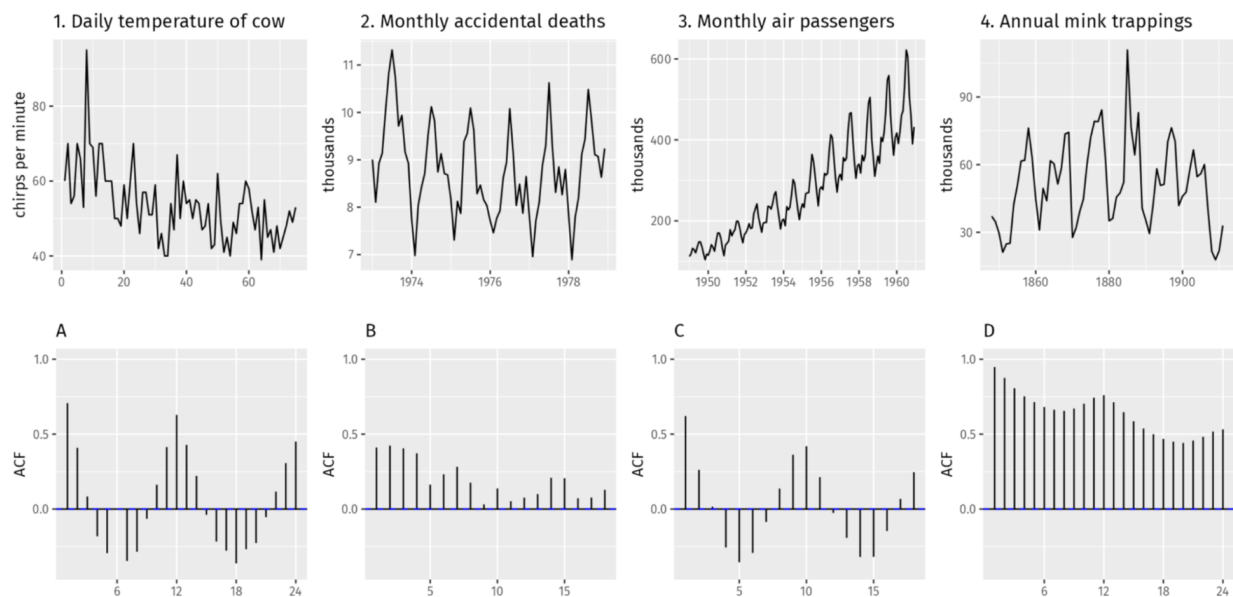
```
ggAcf(myts) +  
  ggtitle("ACF Plot")
```



En este gráfico se confirma lo que se pensaba de la autocorrelación del gráfico anterior.

1.2. Exercise 8

The following time plots and ACF plots correspond to four different time series. Your task is to match each time plot in the first row with one of the ACF plots in the second row.



1.2.1. Solución

Time Series	ACF Plot
1	B
2	A
3	D
4	C

El primer gráfico no presenta tendencia ni estacionalidad de forma clara, debido a que tiene un pico al inicio, aquí se puede presentar una autocorrelación mayor, por lo tanto, el gráfico que más se ajusta a esteo es el **B**. El segundo gráfico de muertes accidentales mensuales presenta un patrón claro de estacionalidad debido a que ocurre el período parece ser de dos años, como muestran una varianza constante, se dice que es la gráfica **A**. En el tercer gráfico hay una tendencia creciente y una estacionalidad con un período de dos años, la intensidad aumenta con el tiempo, por lo tanto tiene una autocorrelación decreciente, la que se asemeja es la figura **D**. Por último, el cuarto presenta un gráfico similar al segundo pero parece ser un comportamiento cíclico en lugar de estacional, se acomoda a la gráfica **C**, mostrando una autocorrelación alternante en el tiempo.

1.3. Exercise 10

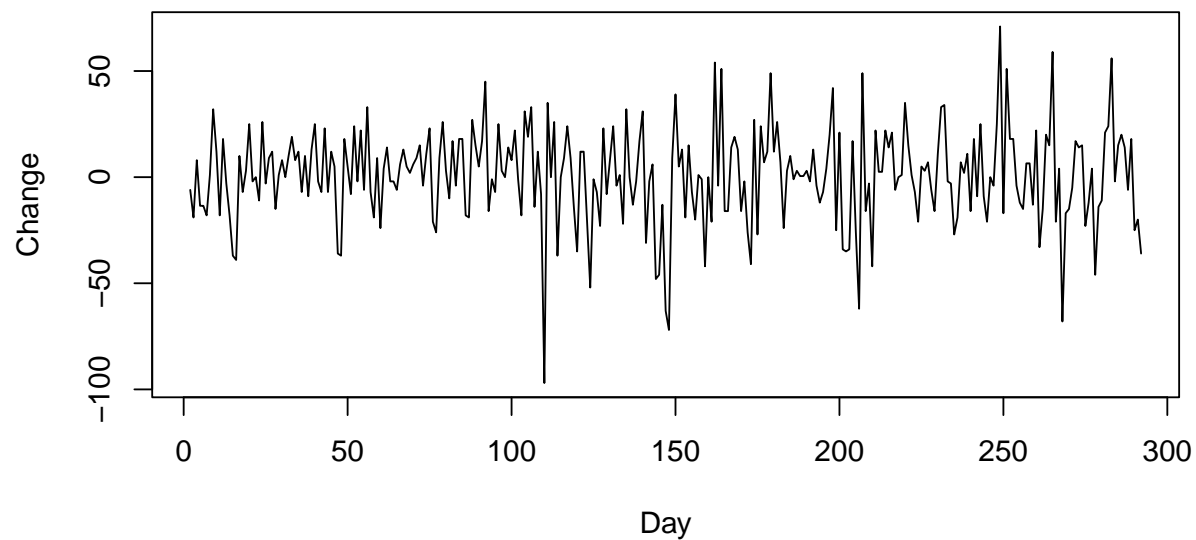
`dj` contains 292 consecutive trading days of the Dow Jones Index. Use `ddj <- diff(dj)` to compute the daily changes in the index. Plot `ddj` and its ACF. Do the changes in the Dow Jones Index look like white noise?

1.3.1. Solución

```
ddj <- diff(dj) # Compute daily differences

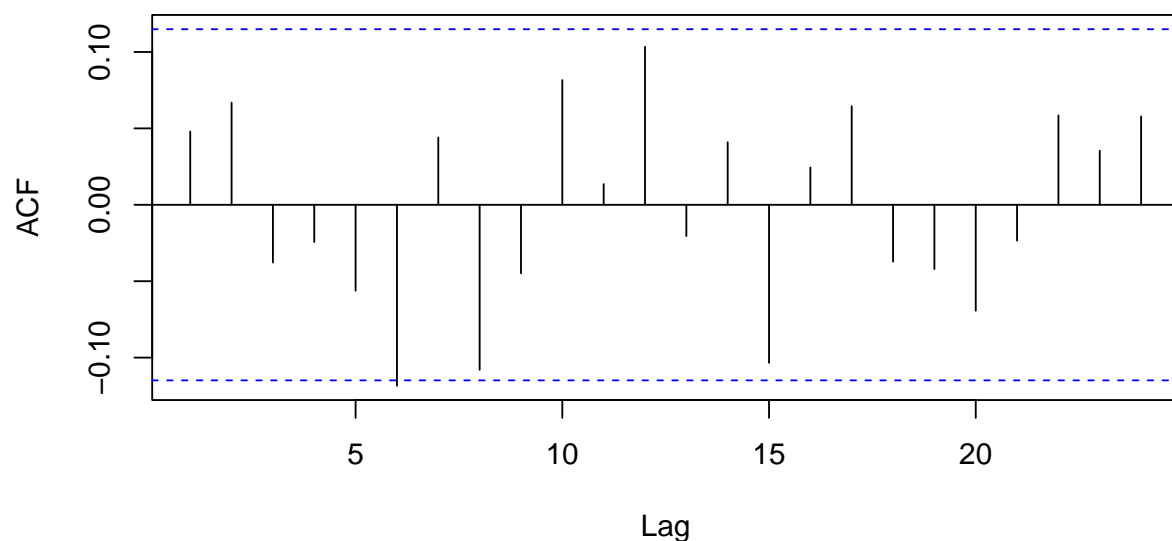
# Plot the time series of differences
plot(ddj, type = "l", main = "Daily Changes in Dow Jones Index",
      ylab = "Change", xlab = "Day")
```

Daily Changes in Dow Jones Index



```
# Plot the ACF  
acf(ddj, main = "ACF of Daily Changes in Dow Jones Index")
```

ACF of Daily Changes in Dow Jones Index



En el primer gráfico se puede observar a simple vista que sí parece ruido blanco. Esto además se confirma en el último gráfico donde no se muestran picos fuera del intervalo de confianza (a excepción de uno al inicio). En conclusión, los cambios en el índice de Dow Jones se pueden ver como ruido blanco.

2. Chapter 3, Section 3.7: Exercises 3, 5, 8

2.1. Exercise 3

What Box-Cox transformation would you select for your retail data (from Exercise 3 in Section 2.10)?

2.1.1. Solución

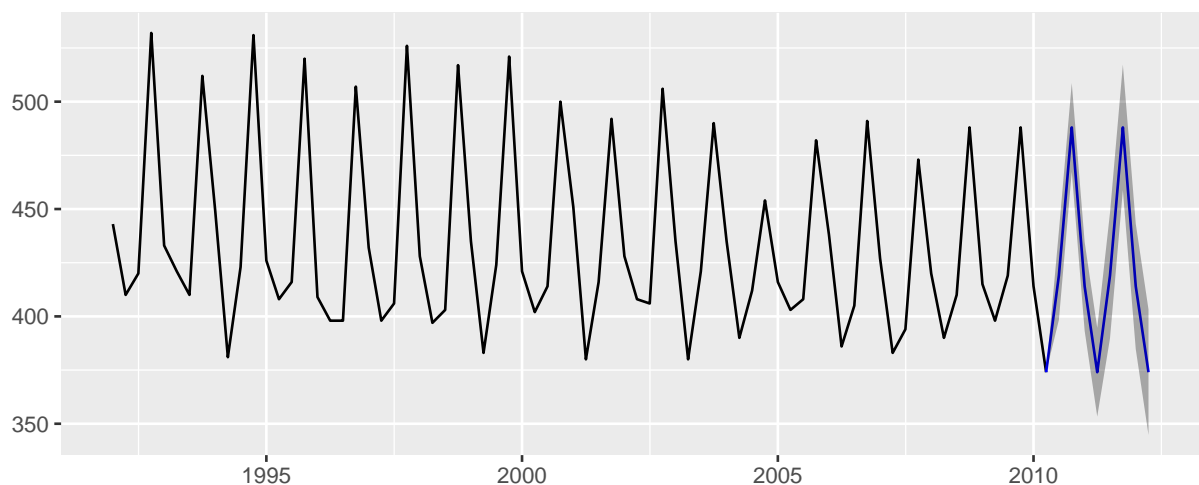
2.2. Exercise 5

Usamos una pronóstico estacional ingenuo (seasonal naïve forecast) sobre los datos de producción de cerveza australiana desde 1992. Luego evaluamos los residuos (diferencia entre los valores reales y los pronosticados) para verificar si:

1. Son ruido blanco (white noise).
2. Están distribuidos normalmente.

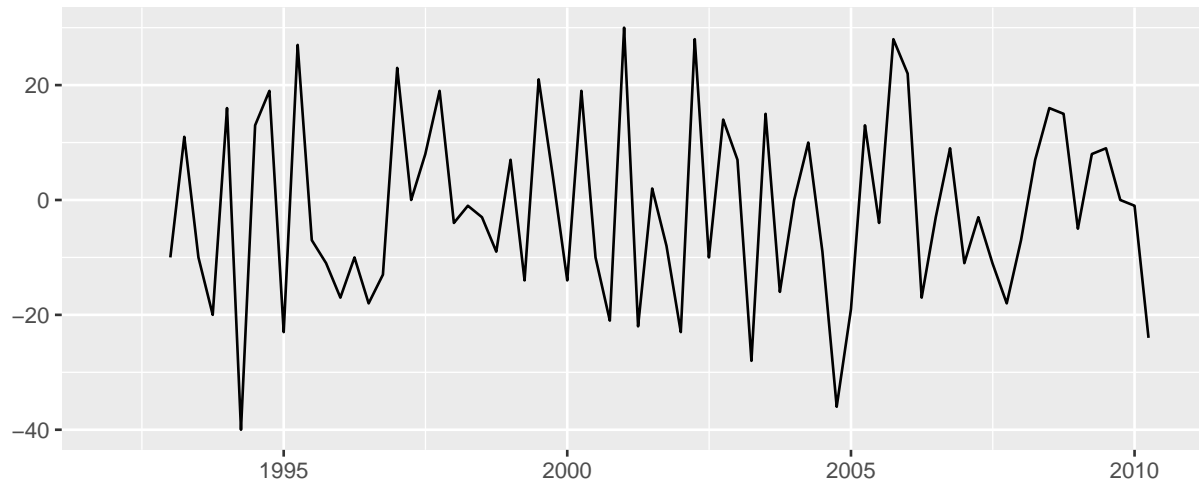
Código proporcionado:

```
beer <- window(ausbeer, start=1992)
fc <- snaive(beer)
autoplot(fc)
```

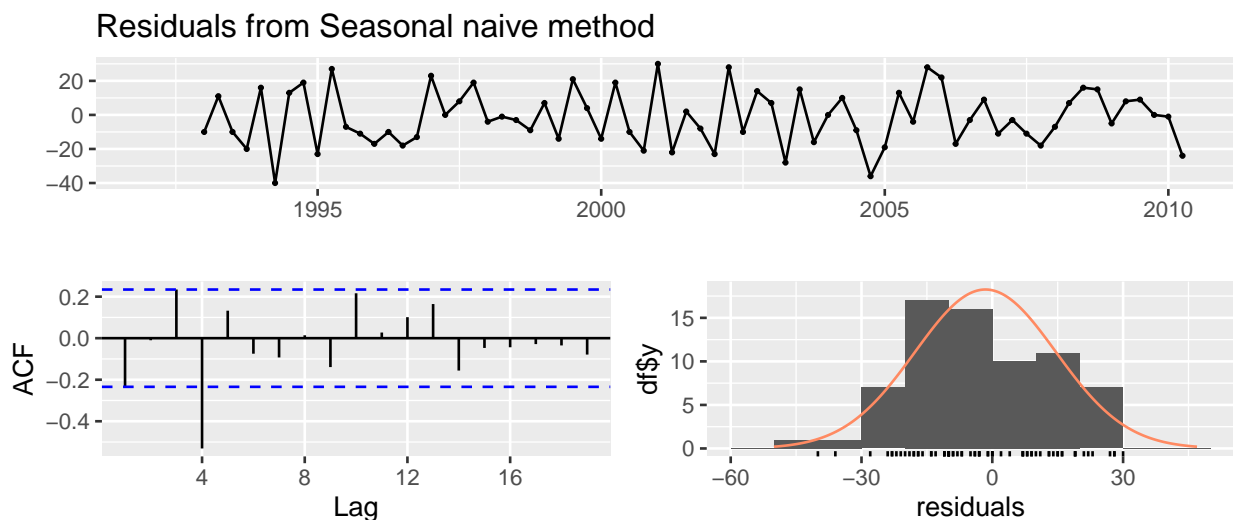


```
res <- residuals(fc)
autoplot(res)
```

```
## Warning: Removed 4 rows containing missing values or values outside the scale
range ('geom_line()').
```



```
checkresiduals(fc)
```



```
##
##  Ljung-Box test
##
## data:  Residuals from Seasonal naive method
## Q* = 32.269, df = 8, p-value = 8.336e-05
##
## Model df: 0.   Total lags used: 8
```

Al aplicar el modelo seasonal naïve a los datos trimestrales de producción de cerveza en Australia desde 1992, los residuos obtenidos no muestran autocorrelación significativa según la prueba de Ljung-Box (valor-p ≥ 0.05), por lo tanto, pueden considerarse como ruido blanco.

En cuanto a la normalidad, los residuos no se desvían drásticamente de una distribución normal, aunque podría haber ligeras desviaciones.

El modelo estacional ingenuo parece ser razonable para estos datos, ya que sus residuos cumplen con las condiciones básicas: ruido blanco y aproximadamente normales.

2.3. Exercise 8

2.3.1. Solución

Para la serie de ventas minoristas correspondiente a la columna A3349337W, se realizó una división de los datos en un conjunto de entrenamiento y otro de prueba de la siguiente manera:

```
myts.train <- window(myts, end=c(2010,12))
myts.test  <- window(myts, start=2011)
```

La visualización combinada de la serie completa junto con las partes de entrenamiento y prueba permite verificar que la división fue realizada correctamente, separando los datos hasta diciembre de 2010 como entrenamiento, y desde enero de 2011 en adelante como prueba.

Posteriormente, se aplicó un modelo de pronóstico estacional ingenuo (snaive) al conjunto de entrenamiento:

```
fc <- snaive(myts.train)
```

Los pronósticos obtenidos se compararon con los valores reales en el conjunto de prueba utilizando la función `accuracy(fc, myts.test)`, lo que permitió calcular métricas como RMSE, MAE y MAPE. Estas métricas ofrecen una evaluación cuantitativa del desempeño del modelo en el período de prueba.

Además, se verificaron los residuos del modelo con la función `checkresiduals(fc)`. En los resultados se analizó:

1. La gráfica de los residuos a lo largo del tiempo.
2. La función de autocorrelación de los residuos (ACF).
3. La prueba de Ljung-Box.

Los residuos no mostraron evidencia clara de autocorrelación significativa, lo que indica que podrían considerarse no correlacionados (ruido blanco). Sin embargo, es posible que no sigan una distribución normal perfecta, aunque suelen aproximarse suficientemente a la normalidad para fines prácticos en este tipo de modelos.

Respecto a la sensibilidad de las métricas de precisión frente a la división de los datos, es importante señalar que los resultados pueden variar dependiendo del período usado para entrenamiento y prueba. Por ejemplo, si el conjunto de prueba contiene eventos atípicos o cambios estructurales (como una crisis económica o una pandemia), los errores de pronóstico pueden aumentar, afectando negativamente las métricas. Por tanto, la evaluación de la precisión de un modelo debe considerar el contexto temporal de la división y, en algunos casos, puede beneficiarse del uso de validación cruzada o múltiples divisiones.

3. Chapter 6, Section 6.9: Exercise 2.

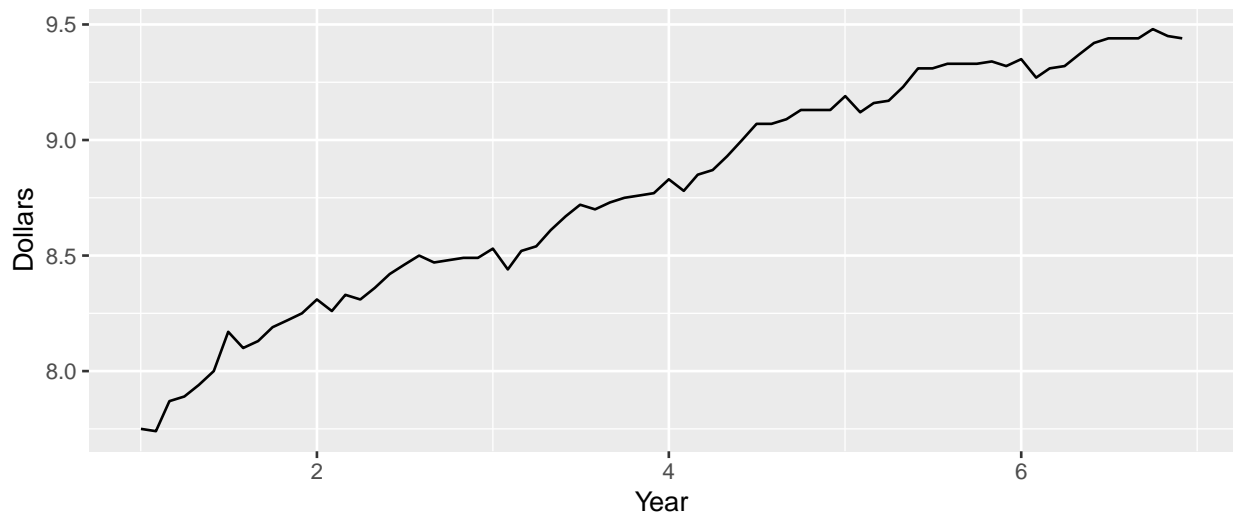
4. Chapter 8, Section 8.11: Exercises 2, 5, 6

5. Exercise 3.5

5.1. Display and interpret the time series plot for these data

5.1.1. Solución

```
data(wages)
wages_ts <- ts(wages, frequency = 12)
autoplot(wages_ts) +
  ylab("Dollars") + xlab("Year")
```



La gráfica muestra una tendencia positiva, indicando que los salarios generalmente han incrementado a lo largo del tiempo. Suelen haber unos picos de caída cada semestre a partir del año 2, sin embargo, son lo suficientemente mínimos para afirmar una posible estacionalidad.

5.2. Use least squares to fit a linear time trend to this time series. Interpret the regression output. Save the standardized residuals from the fit for further analysis.

5.2.1. Solución

```
time <- 1:length(wages_ts)
fit_linear <- lm(wages_ts ~ time)
summary(fit_linear)

##
## Call:
## lm(formula = wages_ts ~ time)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.23828 -0.04981  0.01942  0.05845  0.13136
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.9314358  0.0196657  403.31  <2e-16 ***
## time         0.0234234  0.0004682   50.03  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08257 on 70 degrees of freedom
## Multiple R-squared:  0.9728, Adjusted R-squared:  0.9724
## F-statistic: 2503 on 1 and 70 DF, p-value: < 2.2e-16
```

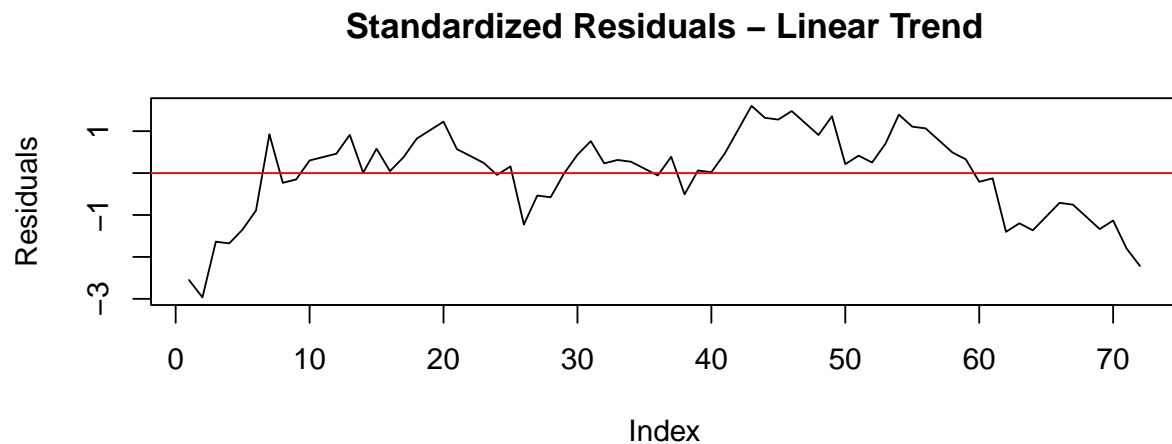
```
resid_linear <- rstandard(fit_linear)
```

Se puede interpretar con respecto al intercepto que es de 7.9314, es decir, al principio del tiempo del estudio los trabajadores ganaban en promedio 7.93 dólares la hora.

5.3. Construct and interpret the time series plot of the standardized residuals from part (b).

5.3.1. Solución

```
plot(resid_linear, type = "l", main = "Standardized Residuals - Linear Trend", ylab = "R")
abline(h = 0, col = "red")
```



Los valores están alrededor del 0, es decir, el modelo es apropiado, aunque no lo explique por completo.

5.4. Use least squares to fit a quadratic time trend to the wages time series. Interpret the regression output. Save the standardized residuals from the fit for further analysis.

5.4.1. Solución

```
fit_quad <- lm(wages_ts ~ time + I(time^2))
summary(fit_quad)

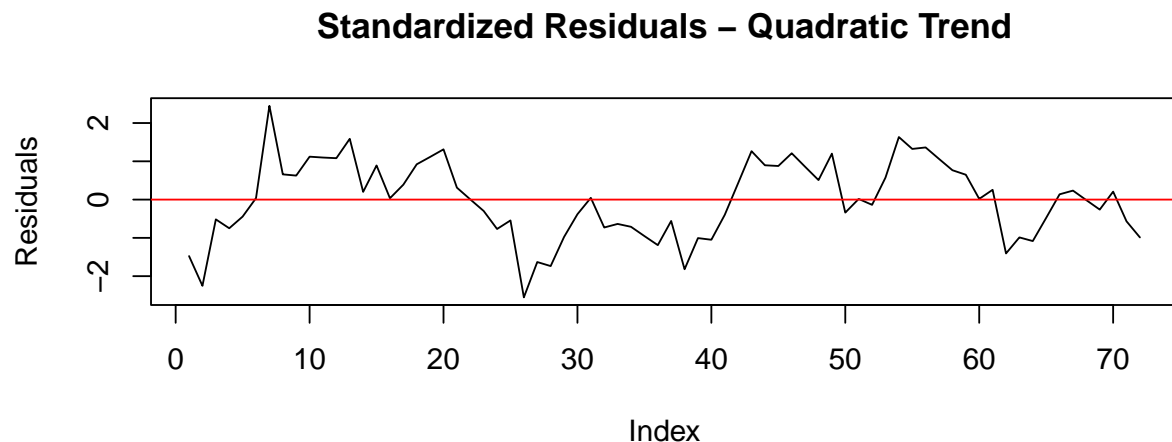
##
## Call:
## lm(formula = wages_ts ~ time + I(time^2))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.148318 -0.041440  0.001563  0.050089  0.139839
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  7.797e+00  2.141e-02 364.127 < 2e-16 ***
## time         3.429e-02  1.354e-03  25.328 < 2e-16 ***
## I(time^2)    -1.488e-04  1.797e-05  -8.282 6.1e-12 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.05889 on 69 degrees of freedom
```

```
## Multiple R-squared:  0.9864, Adjusted R-squared:  0.986  
## F-statistic:  2494 on 2 and 69 DF,  p-value: < 2.2e-16  
  
resid_quad <- rstandard(fit_quad)
```

Ahora se tiene un término adicional al cuadrado y se muestra que es significativo con tres (3) asteriscos. Es decir, el modelo no se asemeja nada a una curva cuadrática.

5.5. Construct and interpret the time series plot of the standardized residuals from part (d)

```
plot(resid_quad, type = "l", main = "Standardized Residuals - Quadratic Trend", ylab = "  
abline(h = 0, col = "red")
```



Teniendo en cuenta la gráfica de los residuales, no hay un patrón claro para poder afirmarse que sea homocedasticidad (pero puede ir por ahí), sin embargo, no logra explicar completamente la estructura de los datos a lo largo del tiempo. Quizá, (hipótesis de los autores) pueden ser los picos chiquitos decrecientes que se presentaban semestralmente en los datos.

6. Exercise 3.11

6.1. Consider the residuals from a least squares fit of a quadratic time trend

6.1.1. Solución

Los residuos estandarizados del ajuste cuadrático muestran variaciones alrededor de cero sin una tendencia clara, lo cual sugiere que el modelo logra capturar adecuadamente la

tendencia principal de los datos. A simple vista, no se observa una variación creciente o decreciente sistemática, lo que indica que podría cumplirse la suposición de homocedasticidad (varianza constante de los residuos). No obstante, se deben realizar pruebas adicionales para verificar si los residuos se comportan como ruido blanco.

6.2. Perform a runs test on the standardized residuals and interpret the results

6.2.1. Solución

Debido a que los residuos no son variables categóricas se van a transformar para que queden valores de 1 u -1, así se puede aplicar la función `runs.test()` para correr los *tests*.

```
# Convertir residuos estandarizados en secuencia de signos: +1 si >0, -1 si <0
signs <- ifelse(resid_quad > 0, 1, -1)
signs_factor <- factor(signs)
runs.test(signs_factor)

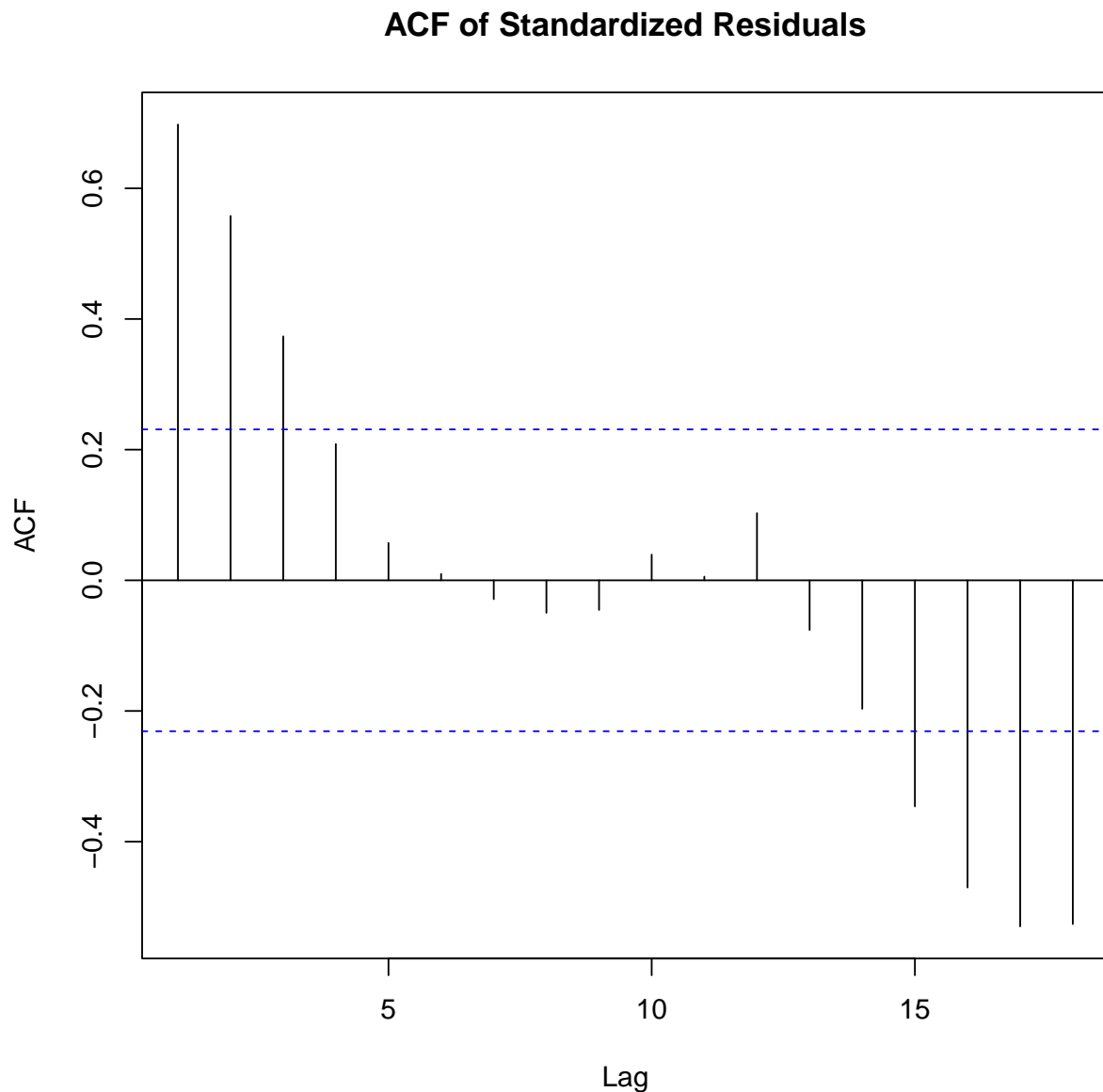
##
##  Runs Test
##
## data:  signs_factor
## Standard Normal = -5.1996, p-value = 1.997e-07
## alternative hypothesis: two.sided
```

Utilizando prueba de rechaz (run test) se evalúa si la serie de tiempo de residuos presenta aleatoriedad. El valor de p es bajo (1.997e-07), esto da a que los residuos no son aleatorios. Por lo tanto, hay un patrón en los signos de los residuos (que fueron categorizados anteriormente).

6.3. Calculate and interpret the sample autocorrelations for the standardized residuals

6.3.1. Solución

```
acf(resid_quad, main = "ACF of Standardized Residuals")
```

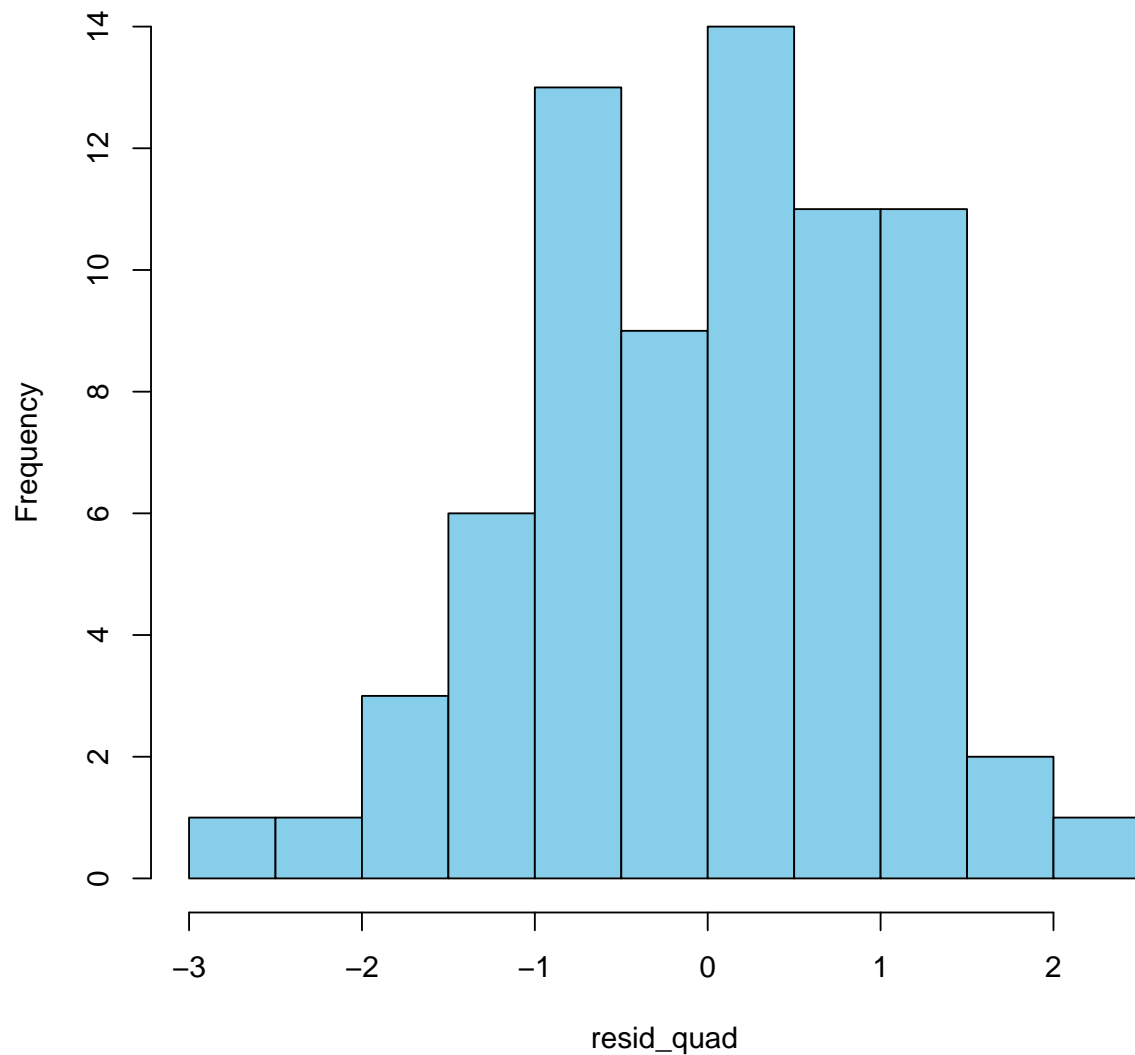
Vemos que en los extremos de los rezagos hay una autocorrelación fuerte, mientras que entre ellos este no logra tener tanta autocorrelación y se encuentra dentro del intervalo de confianza. Es decir, algunos rezagos el modelo logran capturar la estructura de la serie, pero en la mayoría se evidencia que no.

6.4. Investigate the normality of the standardized residuals. Consider histograms and normal probability plots. Interpret the plots.

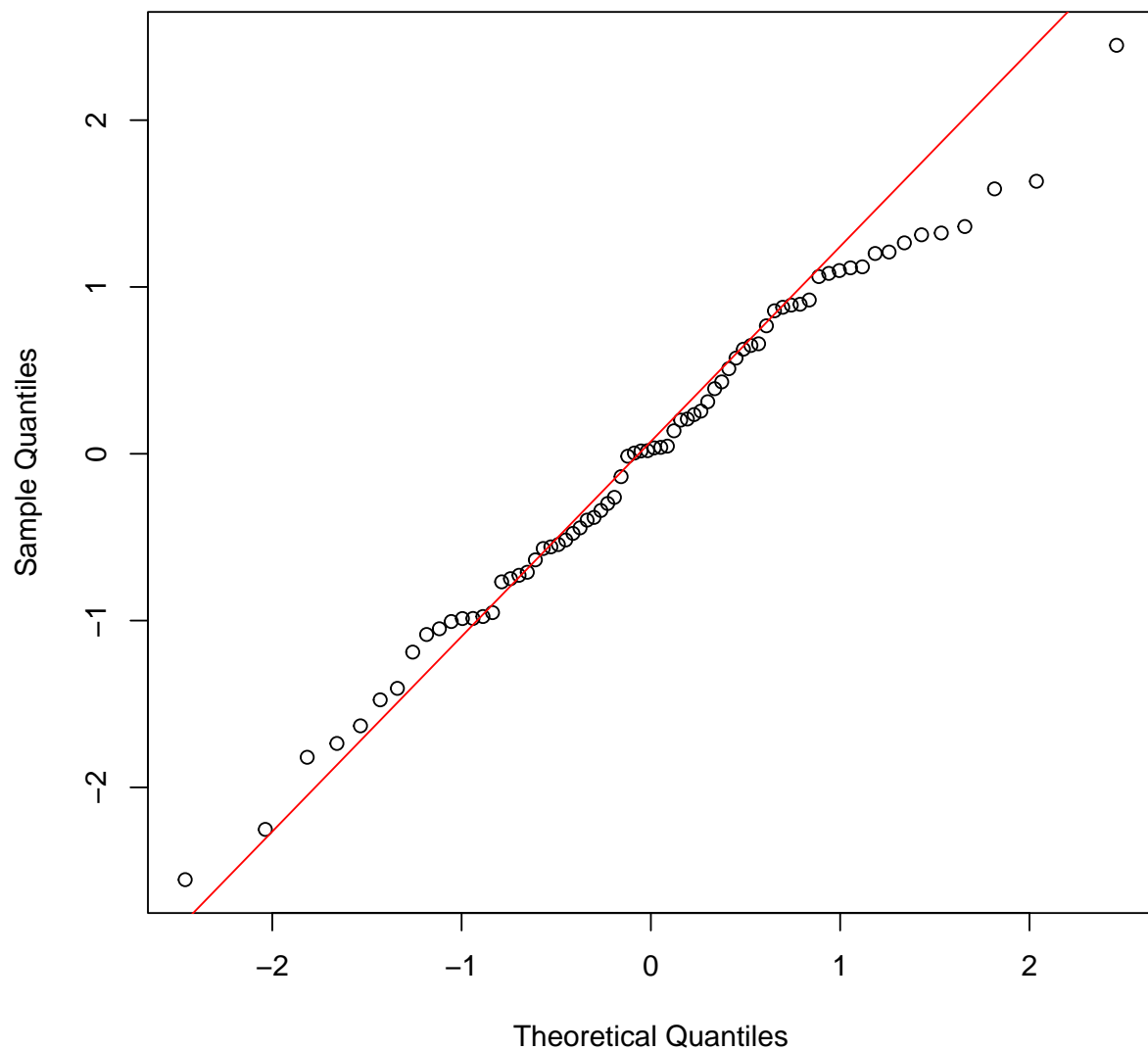
6.4.1. Solución

```
hist(resid_quad, breaks = 10, main = "Histograma de residuos estandarizados", col = "sky
```

Histograma de residuos estandarizados



```
qqnorm(resid_quad, main = "Gráfico Q-Q de residuos estandarizados")  
qqline(resid_quad, col = "red")
```

Gráfico Q-Q de residuos estandarizados

El histograma muestra una simetría aproximada, lo que puede indicar una similitud a la distribución normal. Por otra parte, el gráfico Q-Q muestra que los datos encajan dentro de la línea central. Ambas gráficas juntas sugieren que los residuos siguen una distribución aproximadamente normal.