

# Series de Tiempo 2018

Maestría en Estadística Aplicada, UNR  
Unidad 5

---

Luis Damiano

[damiano.luis@gmail.com](mailto:damiano.luis@gmail.com)

2018-05-10

- Análisis exploratorio
- Descomposición
  - Medias móviles
  - LOESS
  - Regresión lineal
  - Diferenciación
- Ejercicio: Ventas en supermercados

# Estacionalidad

---

# Descomposición de series<sup>1</sup>

## ■ Elementos:

- $\mathcal{T}$  Tendencia: Dirección de largo plazo.
- $\mathcal{S}$  Estacionalidad: Patrón que se repite con una periodicidad conocida.
- $\mathcal{C}$  Ciclo: Patrón que se repite con una periodicidad desconocida y cambiante.
- $\mathcal{E}$  Irregular: Parte no predecible (también conocido como residuo, o error).

## ■ Combinación:

- $\mathcal{N}$  Ninguno: El componente no existe o no es tenido en cuenta.
- $\mathcal{A}$  Aditiva: Suma de componentes.
- $\mathcal{M}$  Multiplicativa: Producto de componentes.
- Pueden aplicarse diferentes operadores para diferentes elementos.

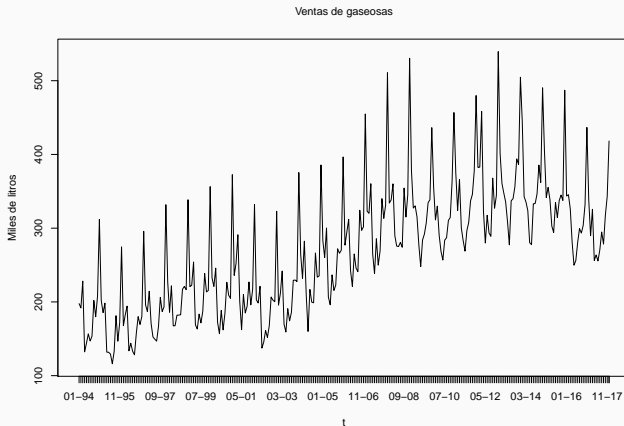
---

<sup>1</sup>R. Hyndman et al. (2008)

Cuando existen tendencia y estacionalidad, suelen aplicarse dos enfoques:

- Modelos clásicos de descomposición.
  - Mínimos cuadrados.
  - Suavizados exponenciales.
- Eliminación por diferenciación.
  - SARIMA.

# Ejemplo

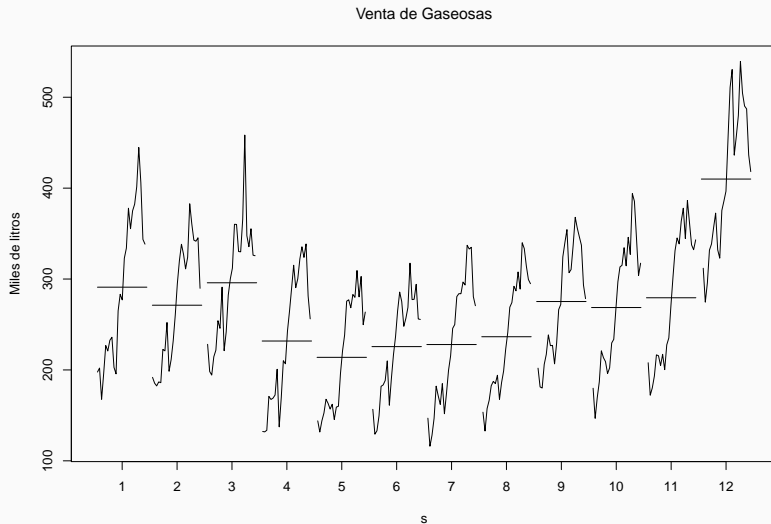


## Discusión en clases

¿Observan un patrón estacional? ¿Cómo lo describirían a simple vista? ¿De qué forma intentarían capturarlo?

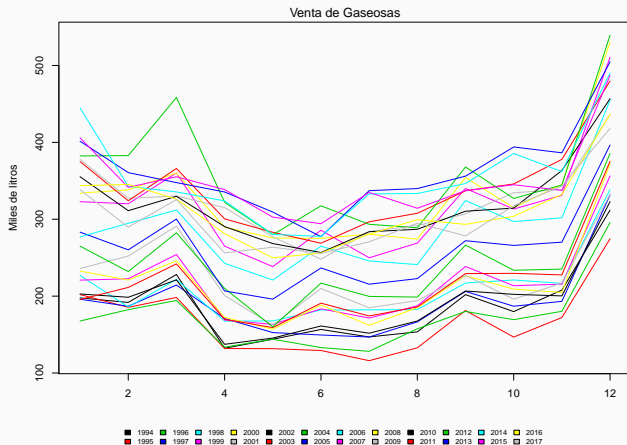
# Tabla de Buys-Ballot

	1	2	3	4	5	6	7	8	9	10	11	12	Media
1994	198	192	228	132	144	157	147	154	202	180	208	312	188
1995	202	185	198	132	132	129	116	133	181	147	172	275	167
1996	168	182	194	134	144	133	128	157	180	169	180	296	172
1997	196	187	214	171	153	149	147	167	206	187	193	332	192
1998	227	186	222	168	168	182	182	183	217	221	217	339	209
1999	221	223	254	169	163	183	172	187	239	214	216	356	216
2000	233	221	246	172	157	188	162	185	227	209	205	373	215
2001	236	252	291	201	162	210	185	194	227	196	217	332	225
2002	203	199	221	137	145	161	152	168	207	202	200	323	193
2003	196	211	242	170	159	191	174	186	229	230	228	375	216
2004	265	232	282	210	160	217	200	199	266	234	235	386	240
2005	283	260	300	207	196	237	216	223	272	266	270	397	261
2006	277	295	312	243	221	265	246	241	324	297	302	455	290
2007	323	320	360	265	238	286	250	269	340	313	331	511	317
2008	334	338	360	290	276	275	281	274	354	315	345	531	331
2009	378	327	330	315	277	248	284	292	307	334	339	436	322
2010	355	311	330	290	268	257	284	287	310	315	363	457	319
2011	375	324	366	300	283	269	297	308	337	346	378	480	339
2012	382	383	458	322	280	317	293	289	368	327	345	539	359
2013	401	361	348	335	309	277	337	340	356	394	386	505	363
2014	445	343	336	324	280	278	333	334	346	386	362	490	355
2015	406	342	355	339	303	294	335	314	337	345	338	487	350
2016	344	345	326	281	250	256	281	299	293	304	332	437	312
2017	338	290	326	256	264	256	271	295	278	317	343	418	304
Media	291	271	296	232	214	226	228	237	275	269	279	410	269





# Gráficos por año



## Discusión en clases

¿Qué conclusiones podrían extraerse sobre la tendencia y la estacionalidad a partir de este gráfico?

# Descomposición con medias móviles

$$Z_t = \mathcal{T} \times \mathcal{S} \times \mathcal{E}$$

- $\mathcal{T}$  Tendencia: Media móvil centrada de 13 elementos.
- $\mathcal{S}$  Estacionalidad: Promedio por mes para todos los años, luego centrado.
- $\mathcal{E}$  Irregular: Diferencia entre el valor observado y la tendencia con estacionalidad ajustada.
- $p \in \mathbb{N}$  períodos completos.

$$\hat{\tau}_t = \frac{1}{24} \left( Z_{t-6} + 2 \sum_{i=-5}^5 Z_{t-i} + Z_{t+6} \right) \quad \hat{s}_j = \frac{\frac{1}{p} \sum_{i=1}^p Z_{ij}}{\frac{1}{T} \sum_{t=1}^T Z_t}, j = 1, 2, \dots, 12 \quad \hat{\varepsilon}_t = \frac{Z_t}{\hat{\tau}_t \times \hat{s}_j}$$

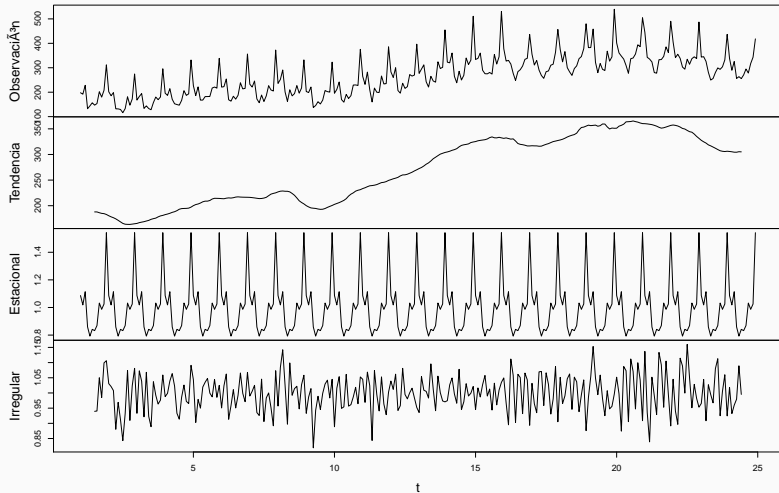
# Descomposición con medias móviles (continuación)

```
z_ts <- ts(z, frequency = 12) # stl requiere un objeto del tipo ts
z_mm <- decompose(z_ts, type = "multiplicative")

print(
  head(z_mm$seasonal, 12),
  digits = 2
)

##      Jan Feb  Mar Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec
## 1  1.09 1.02 1.11 0.86 0.79 0.84 0.83 0.87 1.03 0.99 1.03 1.54
```

# Descomposición con medias móviles (continuación)



# Descomposición con LOESS<sup>2</sup>

$$\ln(Z_t) = \mathcal{T} + \mathcal{S} + \mathcal{E}$$

- $\mathcal{T}$  Tendencia: Polinomio local de grado 1 sobre 13 observaciones móviles.
- $\mathcal{S}$  Estacionalidad: Polinomio local de grado 0 sobre todas las observaciones.
- $\mathcal{E}$  Irregular: Diferencia entre el valor observado y la tendencia con estacionalidad ajustada.

```
z_stl <- stl(log(z_ts[, 1]), t.window = 13, s.window = "periodic")
```

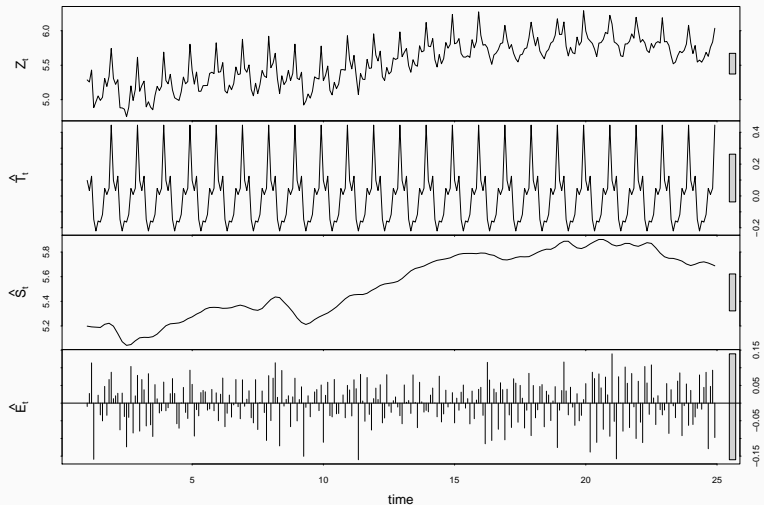
```
print(  
  head(exp(z_stl$time.series[, 1]), 12),  
  digits = 2  
)
```

```
##      Jan  Feb  Mar  Apr  May  Jun  Jul  Aug  Sep  Oct  Nov  Dec  
## 1  1.10  1.03  1.13  0.86  0.80  0.85  0.85  0.89  1.05  1.01  1.05  1.56
```

---

<sup>2</sup>R. J. Hyndman and Athanasopoulos (2018), sec. 6.6. [Ver online.](#)

# Descomposición con LOESS (continuación)



# Descomposición con regresión lineal

$$\ln(Z_t) = \mathcal{T} + \mathcal{S} + \mathcal{E}$$

- $\mathcal{T}$  Tendencia: Regresión lineal por partes con nodos en  $t = 97$  (ene-02) y  $t = 169$  (ene-08).
- $\mathcal{S}$  Estacionalidad: Variable dummy para cada mes (emplea a enero como mes de referencia).
- $\mathcal{E}$  Irregular: Residuos.

```
t1 <- which(index(z) == "2002-01-01")
t2 <- which(index(z) == "2008-01-01")
t   <- 1:length(z)
tdum <- cbind(
  t >= t1 & t < t2,
  t >= t2
) * 1

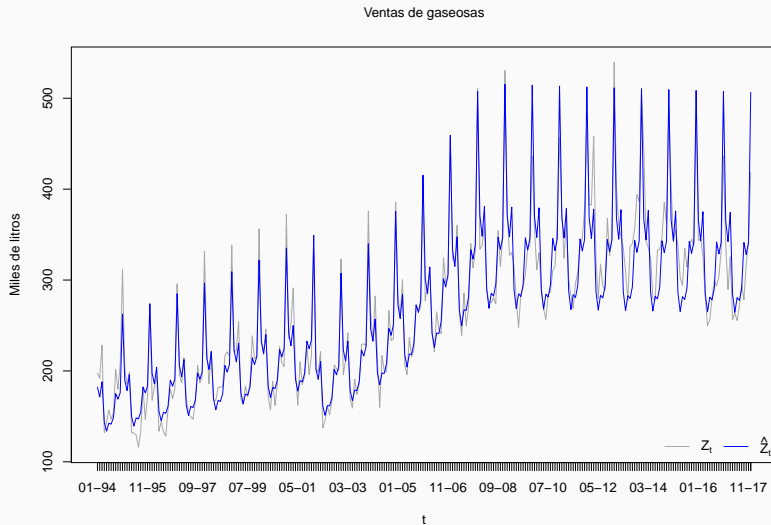
s_t <- factor(format(index(z), "%m"))
fit <- lm(log(z) ~ t * tdum + s_t)
```

# Descomposición con regresión lineal (continuación)

```
##
## Call:
## lm(formula = log(z) ~ t * tdum + s_t)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.238189 -0.048520 -0.001737  0.049714  0.193651
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  5.2006626   0.0233716  222.521 < 2e-16 ***
## t            0.0033899   0.0003083   10.996 < 2e-16 ***
## tdum1       -0.7046228   0.0659050  -10.691 < 2e-16 ***
## tdum2        0.7429329   0.0537105   13.832 < 2e-16 ***
## s_t02       -0.0640867   0.0240995   -2.659  0.00830 **
## s_t03        0.0257131   0.0241015    1.067  0.28698
## s_t04       -0.2472647   0.0241050  -10.258 < 2e-16 ***
## s_t05       -0.3218100   0.0241099  -13.348 < 2e-16 ***
## s_t06       -0.2621003   0.0241161  -10.868 < 2e-16 ***
## s_t07       -0.2726273   0.0241237  -11.301 < 2e-16 ***
## s_t08       -0.2273390   0.0241328   -9.420 < 2e-16 ***
## s_t09       -0.0661341   0.0241432   -2.739  0.00657 **
## s_t10       -0.1052541   0.0241549   -4.357  1.87e-05 ***
## s_t11       -0.0668236   0.0241681   -2.765  0.00608 **
## s_t12        0.3294851   0.0241826   13.625 < 2e-16 ***
## t:tdum1      0.0049683   0.0005647    8.797 < 2e-16 ***
## t:tdum2     -0.0035475   0.0003781   -9.381 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.08348 on 271 degrees of freedom
## Multiple R-squared:  0.9363, Adjusted R-squared:  0.9326
## F-statistic: 249.2 on 16 and 271 DF, p-value: < 2.2e-16
```



# Descomposición con regresión lineal (continuación)



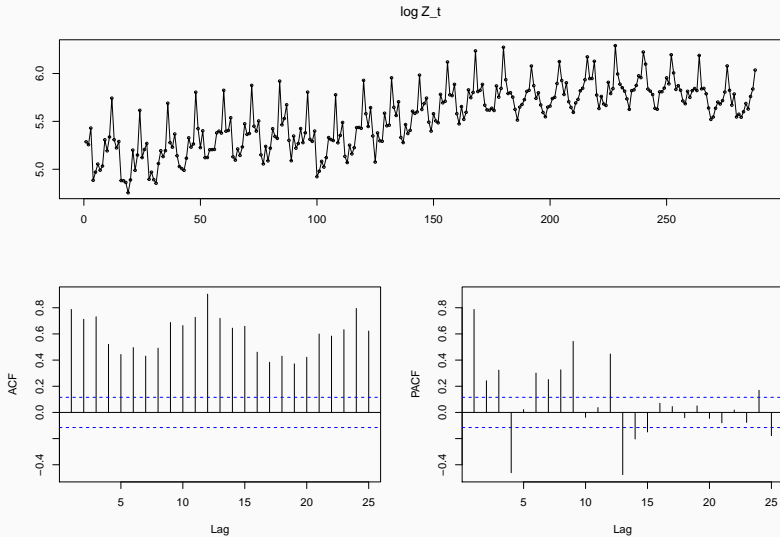
# $SARIMA(1, 1, 1) \times (1, 1, 1)_4$

$$\begin{array}{ccccccc} (1 - \phi_1 B) & (1 - \Phi_1 B^4) & (1 - B) & (1 - B^4) y_t = & (1 + \theta_1 B) & (1 + \Theta_1 B^4) e_t. \\ \uparrow & \uparrow & \uparrow & \uparrow & \uparrow & \uparrow \\ \left( \begin{array}{c} \text{Non-seasonal} \\ \text{AR}(1) \end{array} \right) & & \left( \begin{array}{c} \text{Non-seasonal} \\ \text{difference} \end{array} \right) & & \left( \begin{array}{c} \text{Non-seasonal} \\ \text{MA}(1) \end{array} \right) & & \\ & \uparrow & & \uparrow & & \uparrow & \\ & \left( \begin{array}{c} \text{Seasonal} \\ \text{AR}(1) \end{array} \right) & & \left( \begin{array}{c} \text{Seasonal} \\ \text{difference} \end{array} \right) & & \left( \begin{array}{c} \text{Seasonal} \\ \text{MA}(1) \end{array} \right) & \end{array}$$

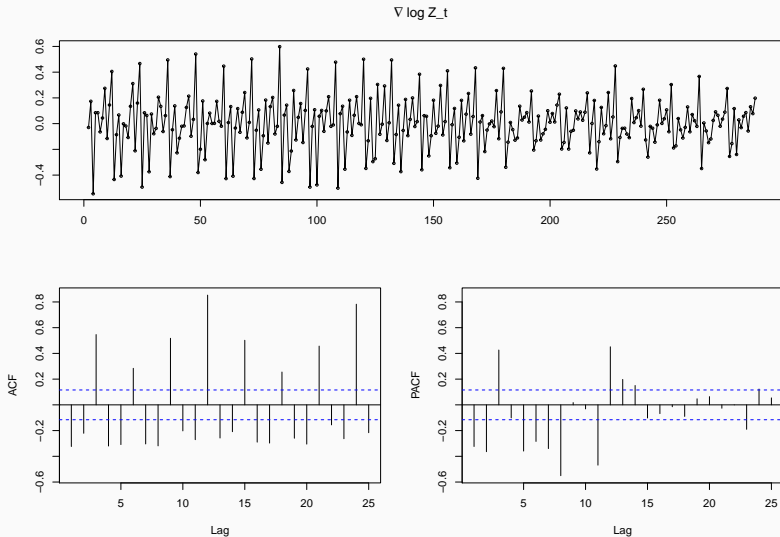
**Figure 1:** R. J. Hyndman and Athanasopoulos (2018). [Ver online](#).

- Transformación logarítmica para estabilizar la varianza.
- Autorregresivo regular de orden  $p = 2$ .
- Diferenciación regular de orden  $d = 1$ .
- Diferenciación estacional de orden  $D = 1$ .
- Media móvil estacional de orden  $Q = 1$ .

# Serie transformada (ln)

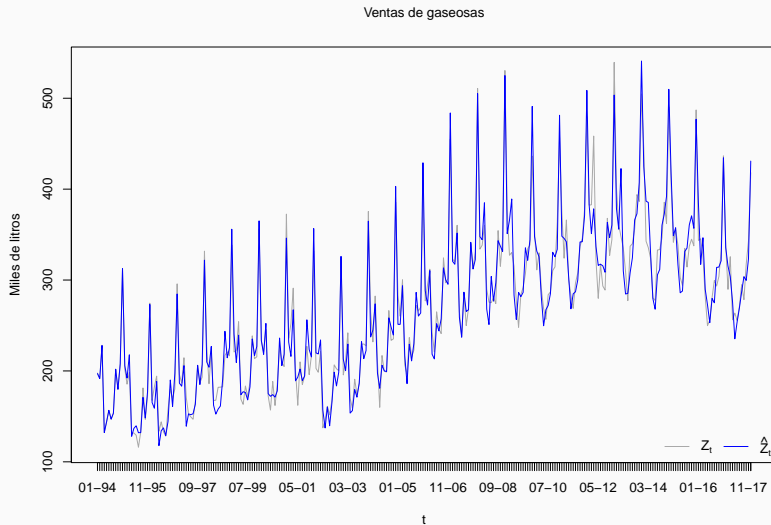


# Primera diferencia de la serie transformada (ln)



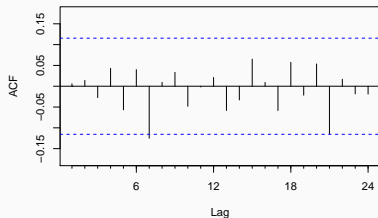
```
fit <- Arima(  
  log(z_ts),  
  order      = c(2, 1, 0),  
  seasonal   = c(0, 1, 1)  
)  
  
print(fit)  
  
## Series: log(z_ts)  
## ARIMA(2,1,0)(0,1,1)[12]  
##  
## Coefficients:  
##          ar1      ar2      sma1  
##      -0.6968  -0.5208  -0.6916  
## s.e.   0.0521   0.0528   0.0503  
##  
## sigma^2 estimated as 0.004267:  log likelihood=357.36  
## AIC=-706.73   AICc=-706.58   BIC=-692.26
```

# Ajuste (continuación)

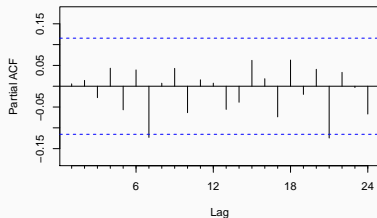


# Diagnóstico de residuos

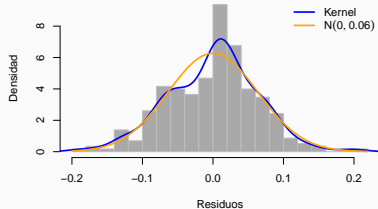
Residuos



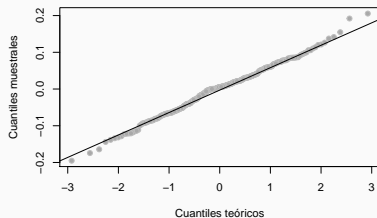
Residuos



Histograma de residuos



QQ Residuos





# Ejercicio: Ventas en supermercados

## Ejercicio en clases

Analizar la serie de tiempo de ejemplo.

Además de ajustar modelos, el ejercicio nos induce a pensar cómo comparar dos modelos (se abordará formalmente en las próximas clases). Solución disponible en el anexo.

Algunos pasos:

- Descargar los datos desde <https://bit.ly/2GXzXoa>.
- De la Sección A 1.11, leer los datos mensuales para la columna *Ventas totales*.<sup>3</sup>
- Graficar y describir la serie original. ¿Es estacionaria en la media y en la varianza? ¿Observan tendencia y estacionalidad? ¿De qué tipo?
- Ajustar un modelo  $SARIMA(0, 1, 3)(1, 0, 2)_{12}$  sobre el logaritmo de las ventas. Además, proponer y ajustar un modelo seleccionado entre las otras metodologías vistas en clases.
- Evaluar los ajustes.
- ¿Cómo evaluarían cuál es mejor? Antes que eso... ¿qué entienden por “mejor”?

---

<sup>3</sup>Hay una copia local en `data/INDECSuper.txt` en caso de que el sitio esté fuera de línea.

## Anexo: Ventas en supermercados

---

# Lectura & procesamiento

```
# https://bit.ly/2GXzXoa
df <- read.table(
  file = "data//INDECSuper.txt",
  header = TRUE,
  sep = "\t"
)

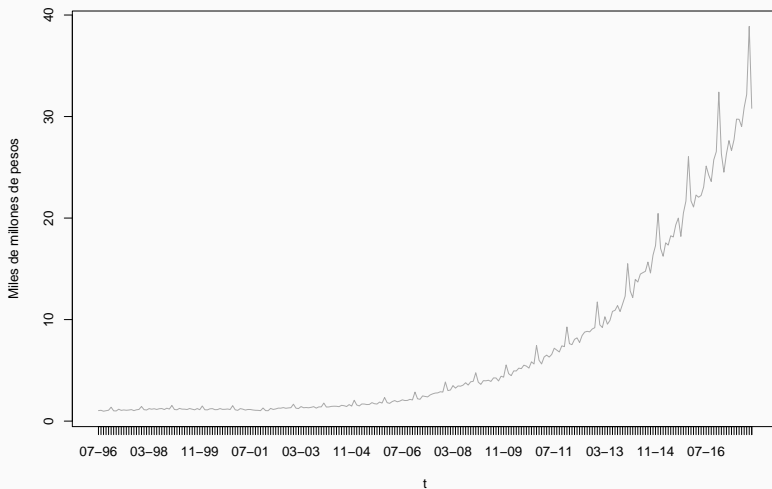
df[, 1] <- as.POSIXct(df[, 1], format = "%Y-%m-%d")

z <- xts(x = df[, 2] / 1000, order.by = df[, 1])
z_ts <- ts(z, frequency = 12) # stl requiere un objeto del tipo ts

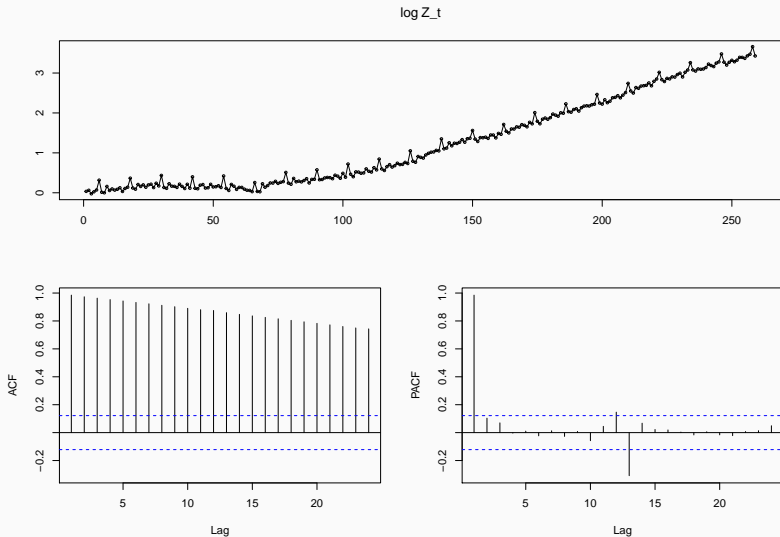
t(head(z, 9))

##      1996-07-01 1996-08-01 1996-09-01 1996-10-01 1996-11-01 1996-12-01
## x           1.036         1.064         0.975         1.025         1.073         1.371
##      1997-01-01 1997-02-01 1997-03-01
## x           1.013           1         1.172
```

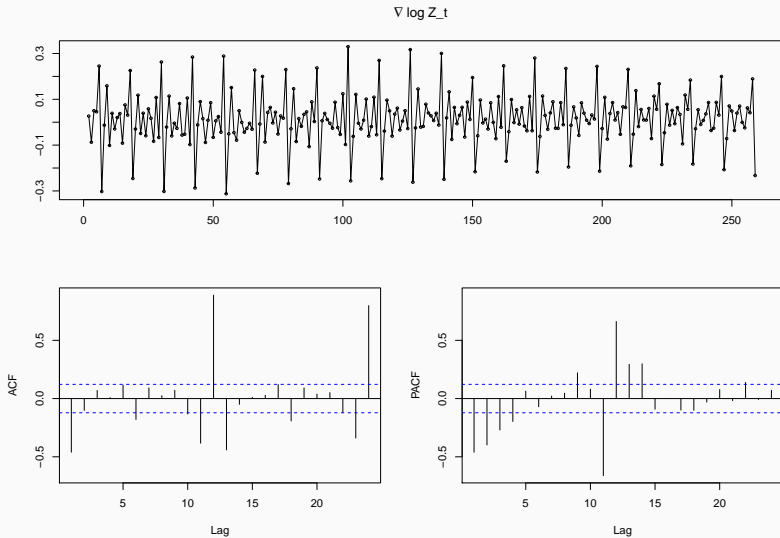
Ventas en supermercados



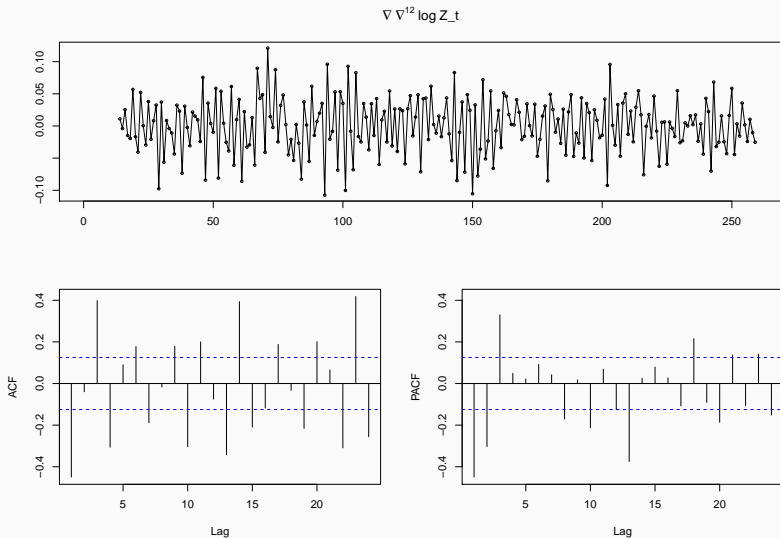
# Serie transformada (ln)



# Primera diferencia de la serie transformada (ln)



# Primera diferencia principal y estacional de la serie transformada (ln)



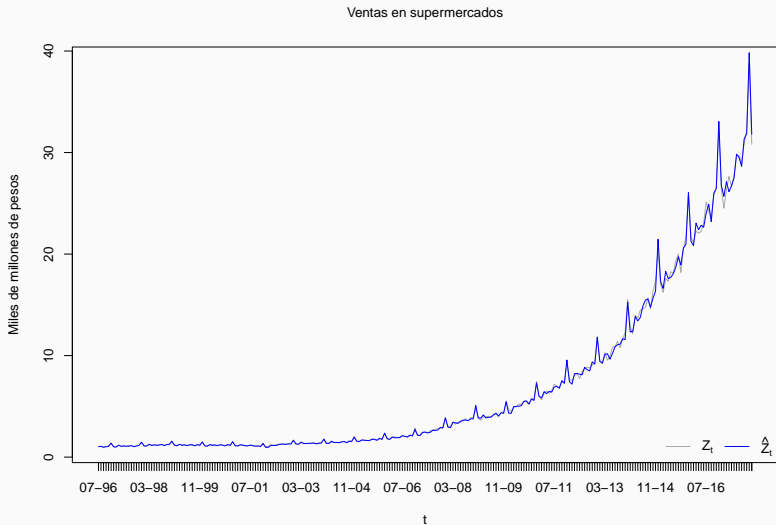
```
fit <- Arima(
  z_ts,
  order      = c(0, 1, 3),
  seasonal   = c(1, 1, 2),
  lambda     = 0
)

print(fit)

## Series: z_ts
## ARIMA(0,1,3)(1,1,2)[12]
## Box Cox transformation: lambda= 0
##
## Coefficients:
##          ma1      ma2      ma3      sar1      sma1      sma2
##      -0.5148  0.0965  0.2535  0.0631  -0.5540  -0.2268
## s.e.   0.0815  0.1102  0.0835  0.1672   0.1587   0.1126
##
## sigma^2 estimated as 0.0009871:  log likelihood=500.43
## AIC=-986.86   AICc=-986.39   BIC=-962.33
```

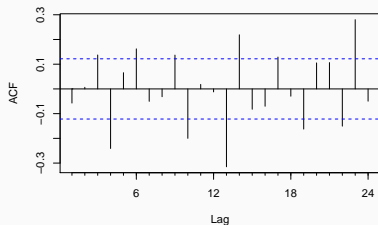


# Ajuste versus observado

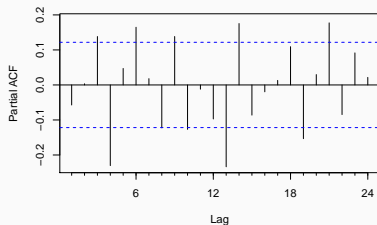


# Diagnóstico de residuos

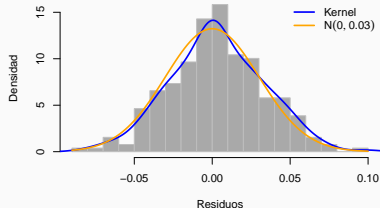
Residuos



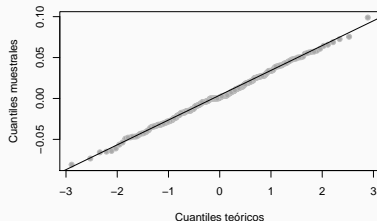
Residuos



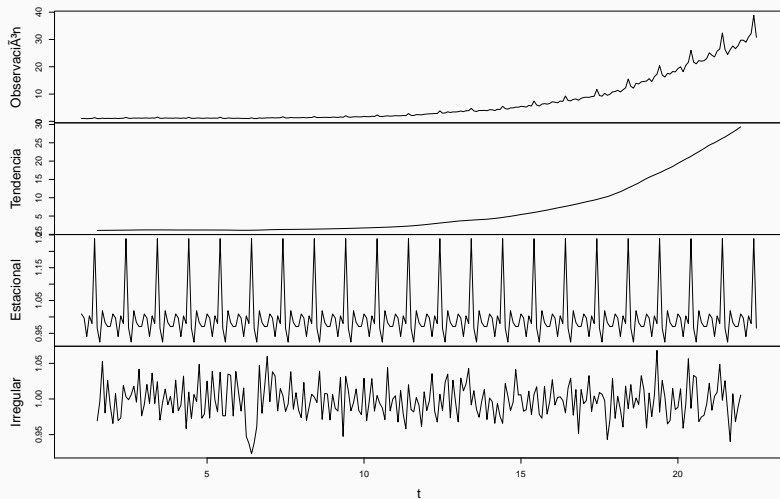
Histograma de residuos



QQ Residuos

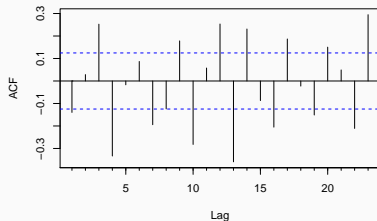


# Descomposición con medias móviles

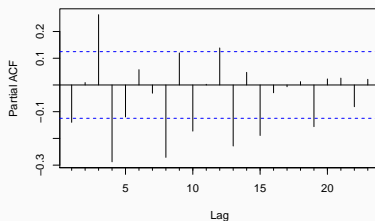


# Componente irregular

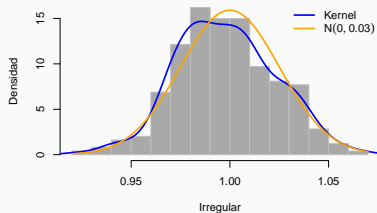
Irregular



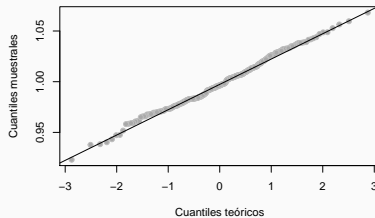
Irregular



Histograma de Irregular



QQ Irregular



Tomar los gráficos sólo a fines descriptivos. La descomposición con medias móviles no hace supuestos distribucionales.

Hyndman, Rob J, and George Athanasopoulos. 2018. *Forecasting: Principles and Practice*.  
<https://otexts.org/fpp2/>.

Hyndman, Rob, Anne Koehler, Keith Ord, and Ralph Snyder. 2008. *Forecasting with Exponential Smoothing*. Springer Berlin Heidelberg. doi:[10.1007/978-3-540-71918-2](https://doi.org/10.1007/978-3-540-71918-2).