

Series de Tiempo 2018

Maestría en Estadística Aplicada, UNR
Unidad 5

Luis Damiano

damiano.luis@gmail.com

2018-04-17

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

Estacionalidad

Descomposición de series¹

■ 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:

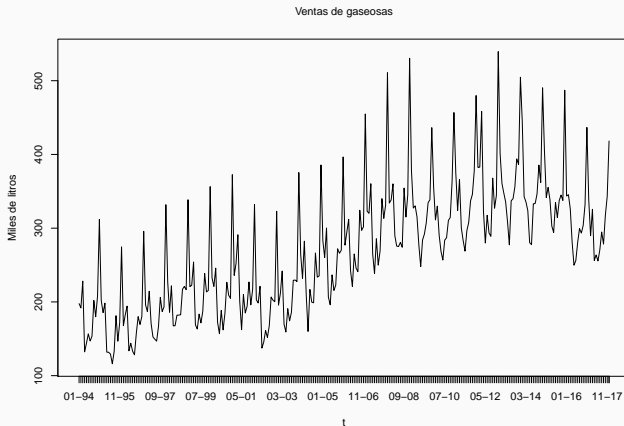
- \mathcal{N} Ninguno: El componente no existe o no es tenido en cuenta.
- \mathcal{A} Aditiva: Suma de componentes.
- \mathcal{M} Multiplicativa: Producto de componentes.
- Pueden aplicarse diferentes operadores para diferentes elementos.

¹R. 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

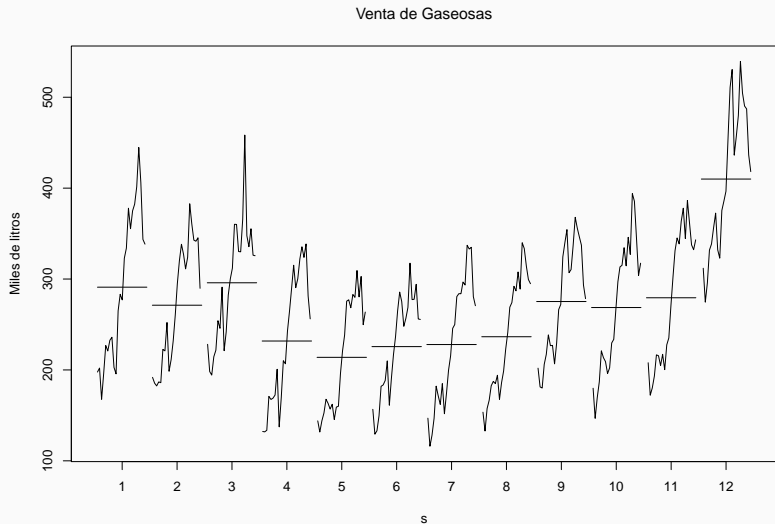


Discusión en clases

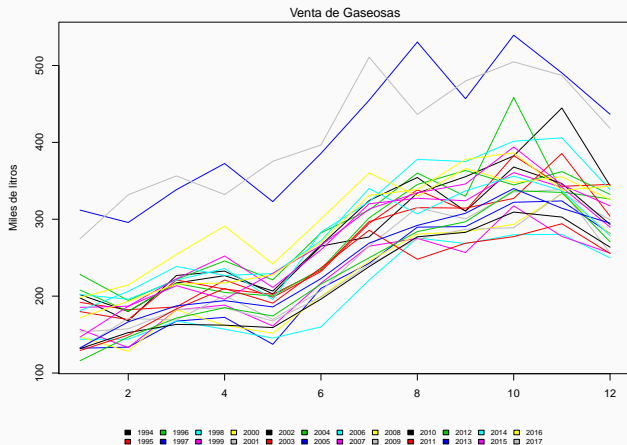
¿Observan un patrón estacional? ¿Cómo lo describirían a simple vista? ¿De qué forma intentarían capturarlo?

Tabla de Buys-Ballot

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

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

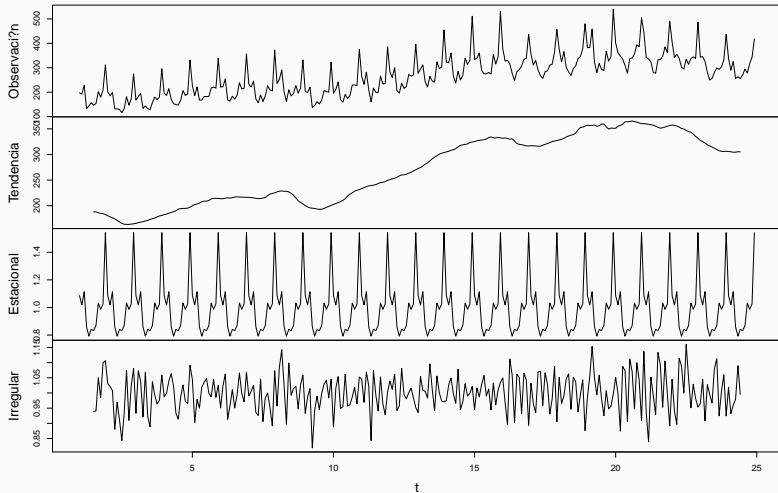
Descomposición con medias móviles (continuación)

```
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 LOESS²

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

- \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.

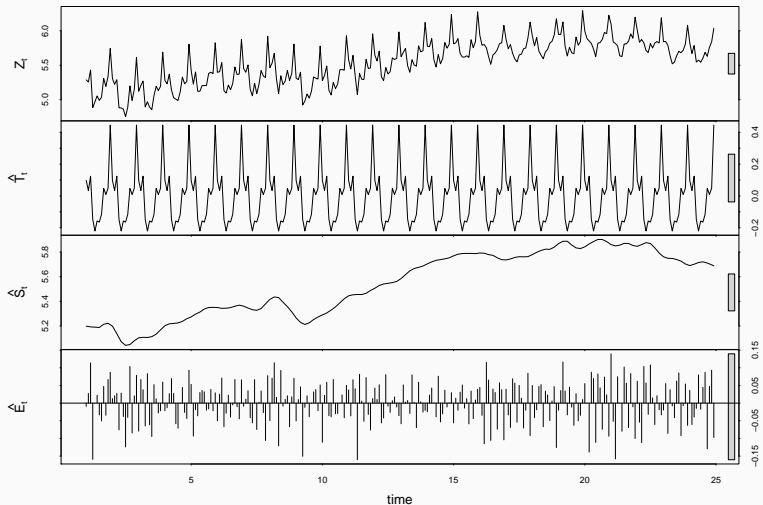
```
z_stl <- stl(log(z_ts[, 1]), t.window = 13, s.window = "periodic")
```

```
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
```

²R. 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.

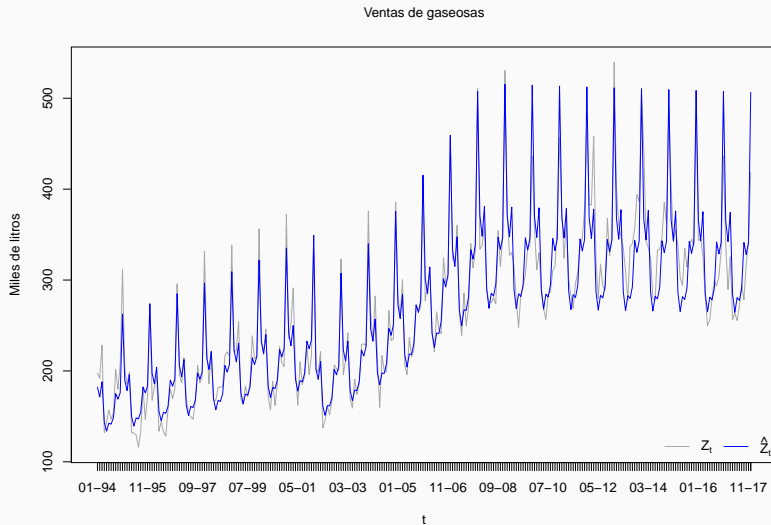
```
t1 <- which(index(z) == "2002-01-01")
t2 <- which(index(z) == "2008-01-01")
t   <- 1:length(z)
tdum <- cbind(
  t >= t1 & t < t2,
  t >= t2
) * 1

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)



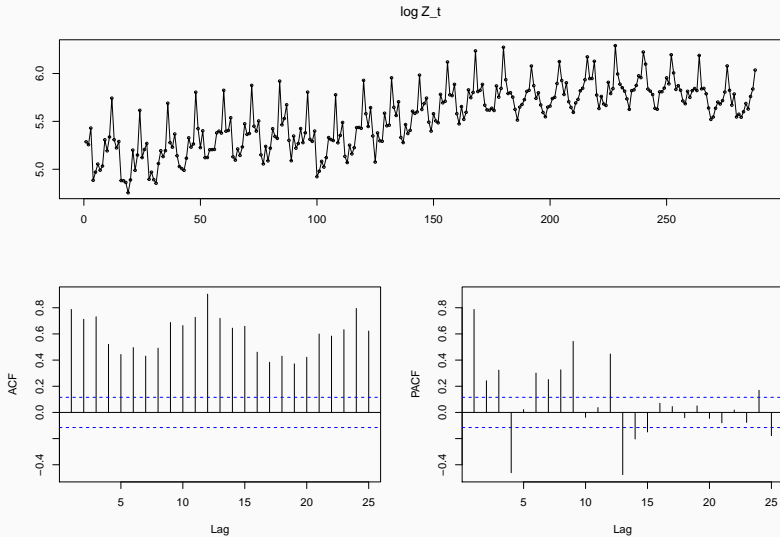
$SARIMA(1, 1, 1) \times (1, 1, 1)_4$

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

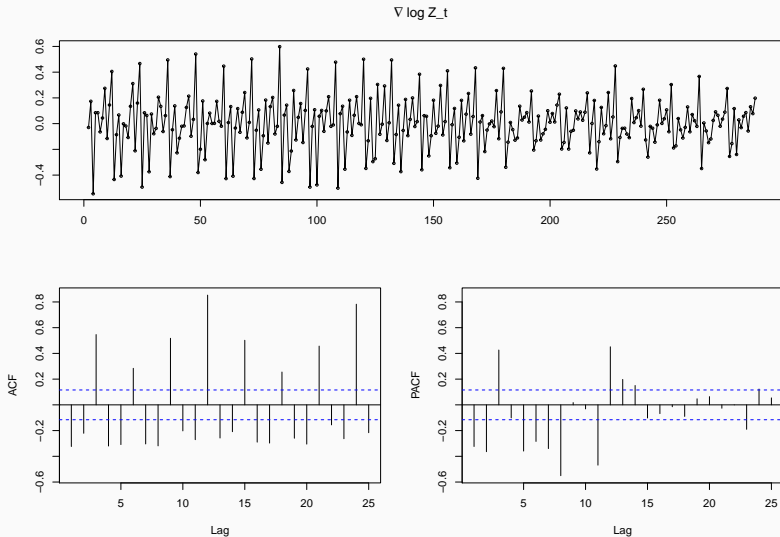
Figure 1: R. J. Hyndman and Athanasopoulos (2018). [Ver online](#).

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

Serie transformada (ln)

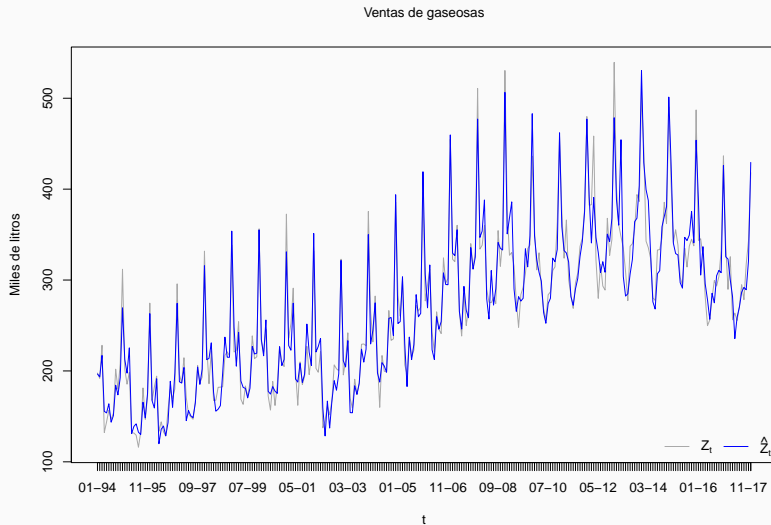


Primera diferencia de la serie transformada (ln)



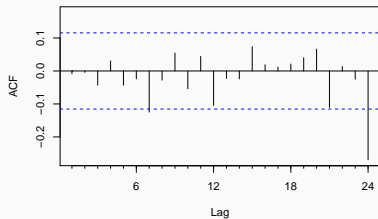
```
fit <- Arima(  
  log(z_ts),  
  order      = c(2, 1, 2),  
  seasonal = c(2, 0, 0)  
)  
  
print(fit)  
  
## Series: log(z_ts)  
## ARIMA(2,1,2)(2,0,0)[12]  
##  
## Coefficients:  
##          ar1      ar2      ma1      ma2      sar1      sar2  
##      -0.7079  -0.5324  0.070  0.0394  0.5360  0.4144  
## s.e.   0.1116   0.0760  0.127  0.1010  0.0555  0.0568  
##  
## sigma^2 estimated as 0.004955:  log likelihood=343.81  
## AIC=-673.62   AICc=-673.22   BIC=-648
```

Ajuste (continuación)

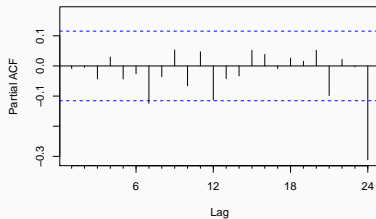


Diagnóstico de residuos

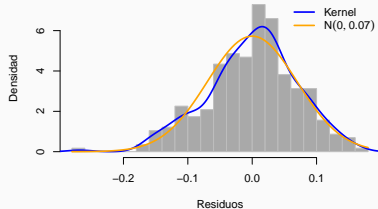
Residuos



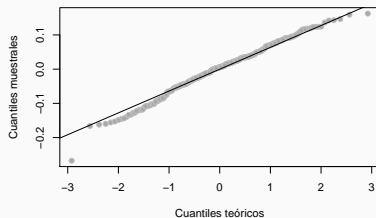
Residuos



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*.³
- 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”?

³Hay 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"
)

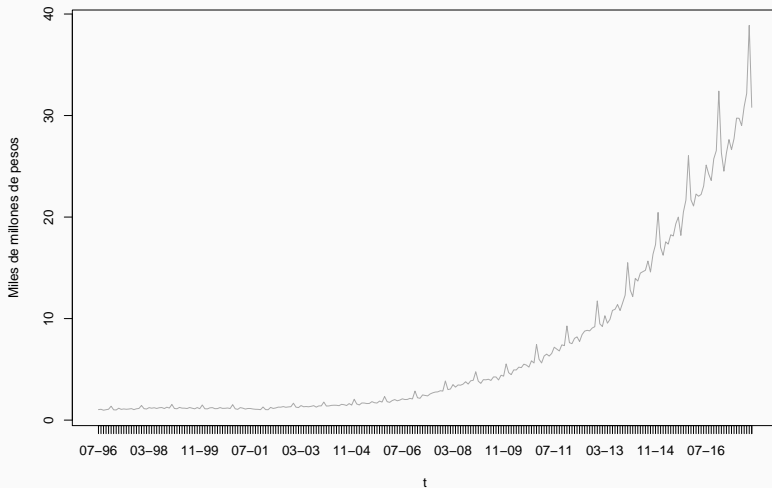
df[, 1] <- as.POSIXct(df[, 1], format = "%Y-%m-%d")

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

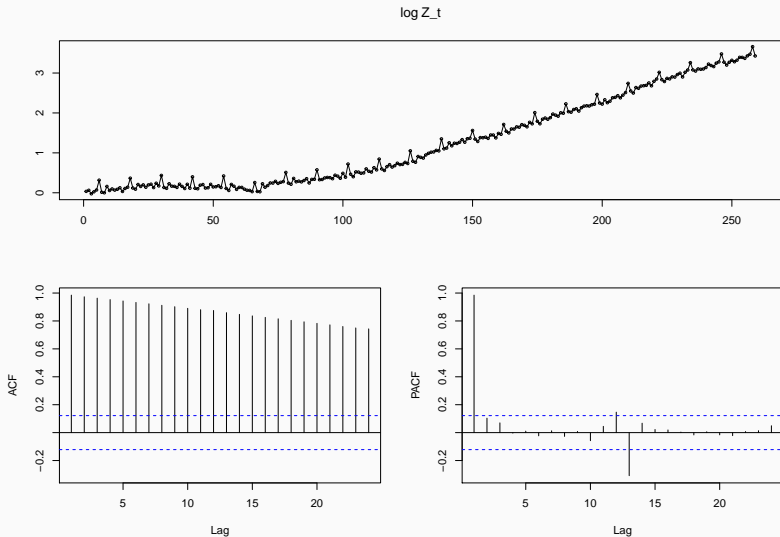
t(head(z, 9))

## 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
```

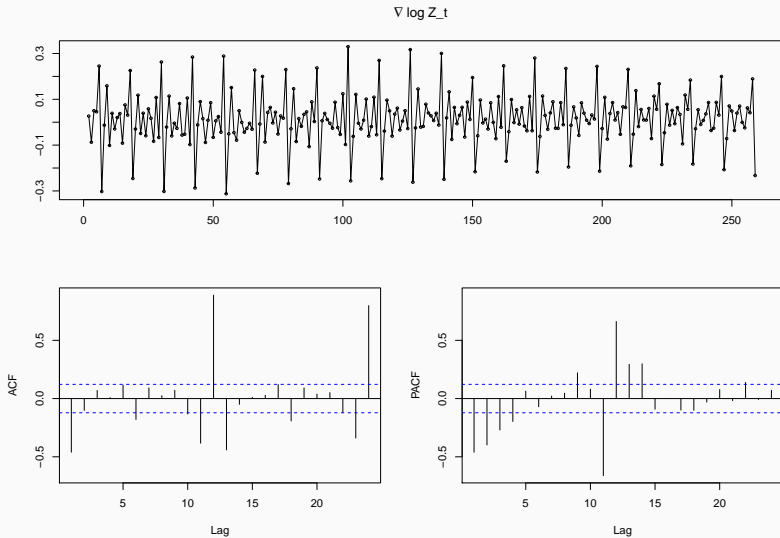
Ventas en supermercados



Serie transformada (ln)

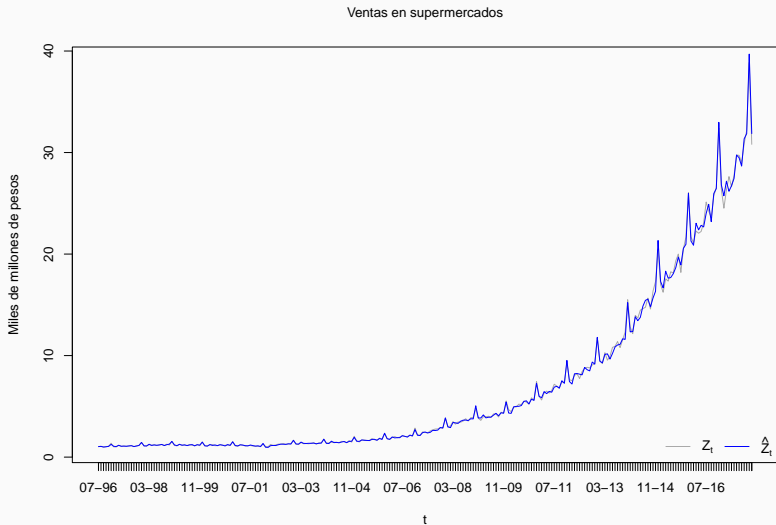


Primera diferencia de la serie transformada (ln)



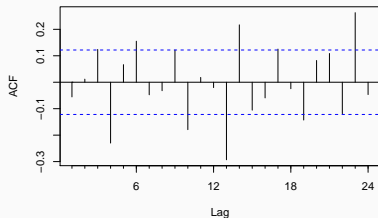
```
fit <- Arima(  
  log(z_ts),  
  order      = c(0, 1, 3),  
  seasonal   = c(1, 0, 2)  
)  
  
print(fit)  
  
## Series: log(z_ts)  
## ARIMA(0,1,3)(1,0,2)[12]  
##  
## Coefficients:  
##          ma1      ma2      ma3      sar1      sma1      sma2  
##      -0.5214  0.0773  0.2583  0.9969  -0.5022  -0.2594  
## s.e.   0.0767  0.1036  0.0789  0.0019   0.0640   0.0607  
##  
## sigma^2 estimated as 0.0009849:  log likelihood=511.67  
## AIC=-1009.33   AICc=-1008.89   BIC=-984.46
```

Ajuste versus observado

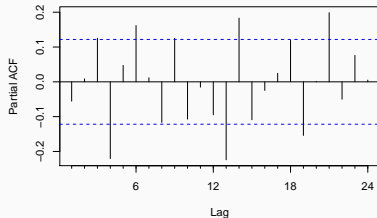


Diagnóstico de residuos

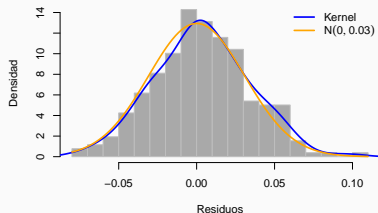
Residuos



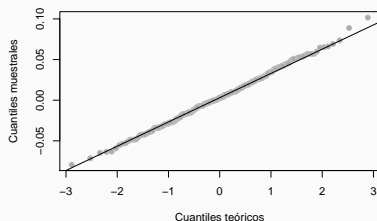
Residuos



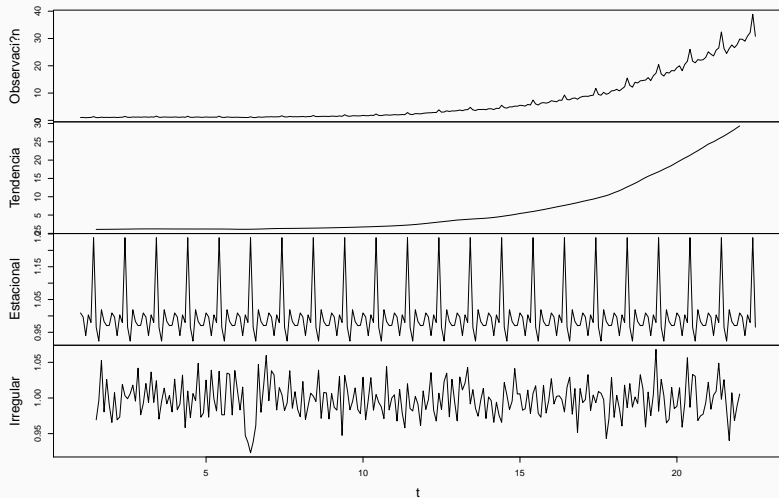
Histograma de residuos



QQ Residuos

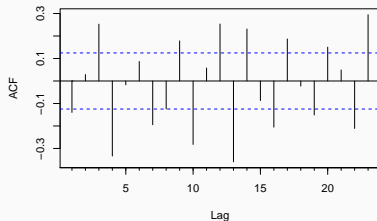


Descomposición con medias móviles

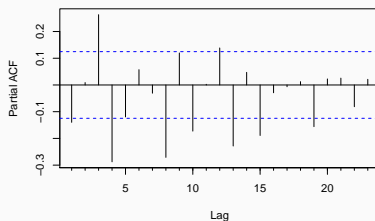


Componente irregular

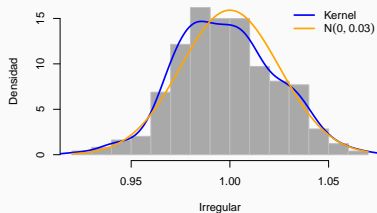
Irregular



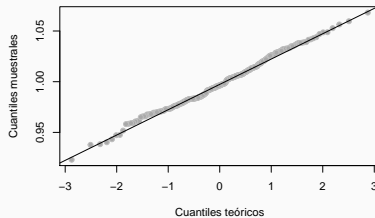
Irregular



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).