Modelos ARIMA (2)

Gustavo A. Caffaro

Instituto Tecnológico de Santo Domingo

Nov 2021 - Ene 2022

Introducción

Contenido:

- Modelos ARIMA con estacionalidad
- Eliminando la estacionalidad
- Selección del modelo

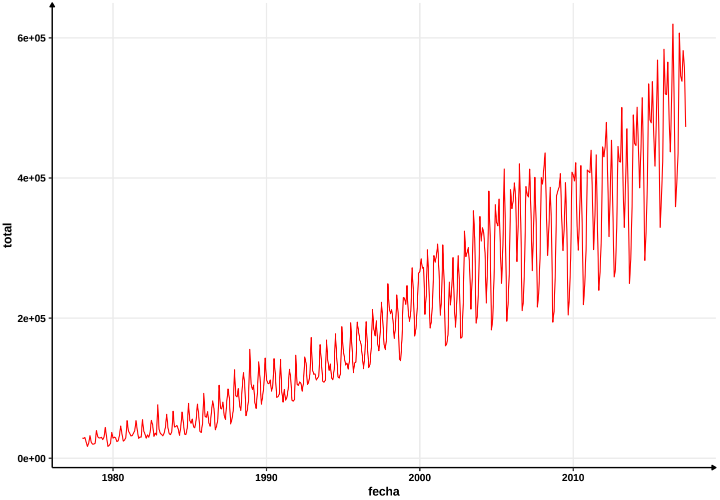
Estacionalidad

La estacionalidad en una serie de tiempo es un patrón que se repite sobre S períodos, donde S define el número de períodos hasta que los patrones vuelvan a repetirse de nuevo.

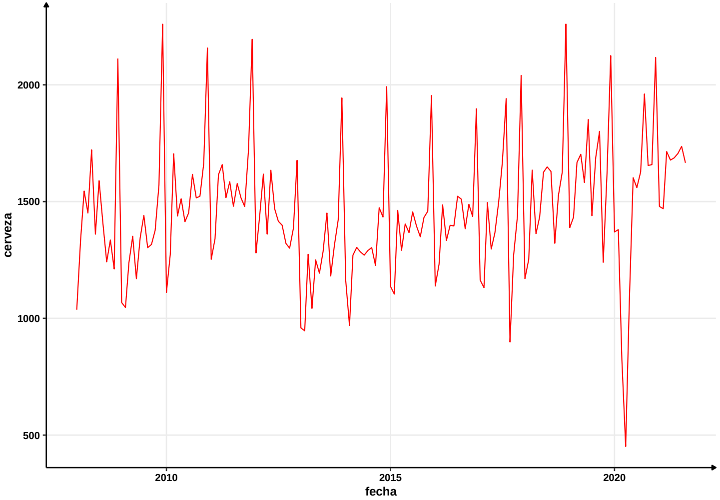
Usualmente,

- Si la data tiene frecuencia mensual, entonces $S = 12$
- Si la data tiene frecuencia, entonces $S = 4$

Llegada de turistas



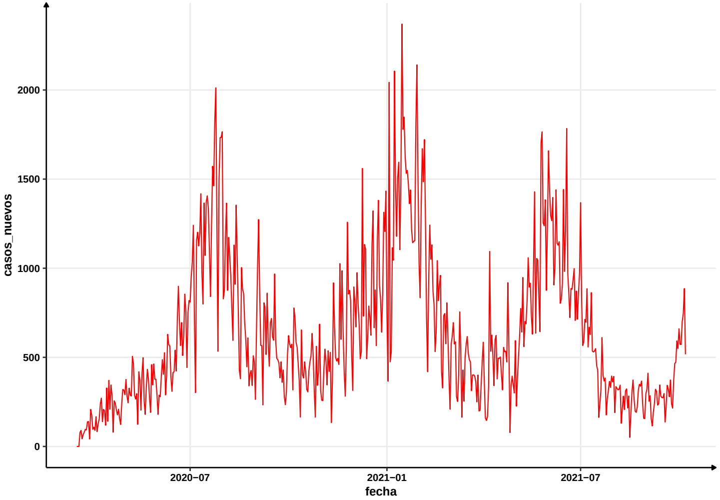
Consumo de Cerveza



Estacionalidad

Si tenemos data con frecuencia diaria, entonces podría tener estacionalidad semanal, por lo que $S = 7$.

Casos de Covid en RD



Modelo ARIMA Estacional

En un modelo estacional ARIMA, componentes estacionales de AR y MA predicen x_t, usando valores y errores con rezagos que son múltiplos de S.

- Con datos mensuales ($S = 12$), un modelo autoregresivo estacional de primer orden usa x_{t-12} para predecir x_t.
  - Ej.: Si quisiéramos proyectar el PIB de diciembre 2021 usuaríamos el PIB de diciembre 2020. Esto aplica para todos los meses del año.
- Un modelo autoregresivo estacional de segundo orden utiliza x_{t-12} y x_{t-24} para predecir x_t.
- Un modelo MA(1) estacional usa w_{t-12} como predictor. Un modelo MA(2) estacional utilizaría $w_1 2$ y $w_2 4$.

Diferenciando

- Casi que por definición, es necesario evaluar series diferenciadas cuando tenemos estacionalidad.
- La estacionalidad usualmente causa que la serie sea no estacionaria, ya que el promedio en algunos períodos puede ser muy diferente al de los demás.

Diferencia estacional

La diferencia estacional se define como la diferencia entre un valor y y su rezago con múltiplo S.

- Si $S = 12$, entonces la diferencia estacional es

$$(1 - L^{12})x_t = x_t - x_{t-12}$$

- Si $S = 4$, la diferencia estacional es

$$(1 - L^{4})x_t = x_t - x_{t-4}$$

- Se puede combinar una diferencia estacional con una no estacional:

$$(1 - L^{12})(1 - L)x_t = (x_t - x_{t-1}) - (x_{t-12} - x_{t-13})$$

El componente no-estacional sigue importando...

Con series que exhiben estacionalidad, es muy posible que componentes no estacionales contribuyan con el modelo.

Modelo ARIMA Estacional (SARIMA)

El modelo ARIMA estacional incorpora tanto factores no estacionales como estacionales en un modelo multiplicativo:

$$\text{ARIMA}(p, d, q) \times (P, D, Q)S,$$

donde

- p es el orden AR no estacional
- d es el orden de integración no estacional
- q es el orden MA no estacional
- P es el orden AR estacional
- D es el orden de integración estacional
- Q es el orden MA estacional
- S es el período estacional

Modelo ARIMA Estacional (SARIMA)

Formalmente, el modelo SARIMA sin diferencias se escribe:

$$\Psi(L^S)\phi(L)(x_t - \mu) = \Theta(L^S)\theta(L)w_t$$

- Los componentes no estacionales son:
  - AR: $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$
  - MA: $\theta(L) = 1 + \theta_1 L + \cdots + \theta_q L^q$
- Los componentes estacionales son:
  - AR estacional: $\Psi(L^S) = 1 - \phi_1 L^S - \cdots - \phi_P L^{PS}$
  - MA estacional: $\Theta(L^S) = 1 + \Theta_1 L^S + \cdots + \Theta_Q L^{QS}$

En el lado izquierdo de la ecuación anterior, los componentes AR estacional y no estacional se multiplican entre sí, mientras que del lado derecho los componentes MA estacionales y no estacionales se multiplican entre sí.

Ej. SARIMA$(0,0,1) \times (0,0,1)_{12}$

Este modelo incluye un componente MA(1) no estacional, uno MA(1) estacional, sin diferencias, ningún componente autoregresivo, y con período estacional $S = 12$.

- El polinomio del componente no estacional MA(1) es $\theta(L) = 1 + \theta_1 L$
- El polinomio del componente estacional MA(1) es $\Theta(L^{12}) = 1 + \Theta_1 L^{12}$

$$(x_t - \mu) = \Theta(L^{12})\theta(L) = (1 + \Theta_1 L^{12})(1 + \theta_1 L)w_t.$$

Resolviendo, tenemos

$$(x_t - \mu) = (1 + \theta_1 L + \Theta_1 L^{12} + \theta_1 \Theta_1 L^{13})w_t$$

y

$$(x_t - \mu) = w_t + \theta_1 w_{t-1} + \Theta_1 w_{t-12} + \theta_1 \Theta_1 w_{t-13}$$

Identificando un modelo estacional

1. Graficar la serie
2. Diferenciar en caso de ser necesario
3. Examinar las funciones de autocorrelación
4. Estimar los modelos sugeridos por el paso 3
5. Examinar los residuales

1. Graficar la serie

Examinar el gráfico de la serie, buscando características como tendencias y estacionalidad. En particular, observar si existen patrones recurrentes dependiendo de la frecuencia colectada.

2. Diferenciar en caso de ser necesario

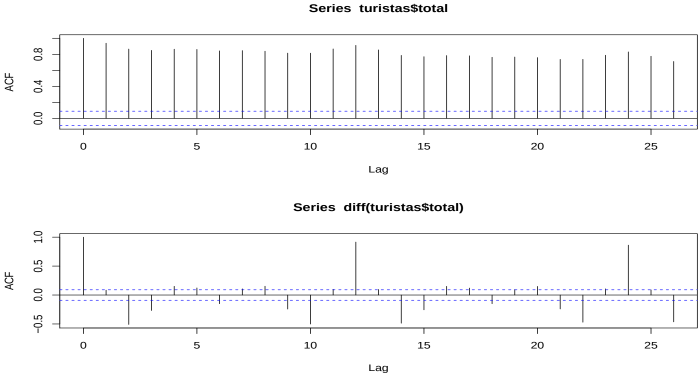
Si la serie exhibe estacionalidad sin tendencia, entonces se puede tomar una diferencia con el rezago S. Por ejemplo, si los datos son mensuales, entonces se puede tomar una diferencia interanual.

Si la serie exhibe tendencia lineal y ninguna estacionalidad obvia, entonces toma la primera diferencia. En caso de que la tendencia no sea lineal, considera transformar la serie antes de diferenciar.

Finalmente, si la serie aparenta tener tanto tendencia como estacionalidad, aplica la diferencia estacional y luego reevalúa la tendencia. Si se mantiene una tendencia, entonces toma la primera diferencia.

3. Examinar las funciones de autocorrelación

La función de autocorrelación de una serie con estacionalidad muestra oscilaciones con picos en los múltiplos de S.



3. Examinar las funciones de autocorrelación

Usualmente, se consideran varios modelos posibles partiendo de la información que muestran estas funciones. En general, usamos esta guía:

- Términos no estacionales: Examina los primeros rezagos para evaluar los términos no estacionales. Picos en la FAC (en los primeros rezagos), en adición a términos de autocorrelación parcial cada vez más estrechos sugieren términos MA no estacionales. En cuanto a términos AR podemos buscar una caída exponencial en los primeros términos, acompañado de picos en el autocorrelación parcial.
- Términos estacionales: Examina los patrones de los rezagos multiplos de S. Por ejemplo, para series mensuales, observa los rezagos 12, 24, 36, y así sucesivamente.

4. y 5. Estimar los modelos y evaluar los residuales

Luego de tener un conjunto de posibles candidatos a partir de los pasos anteriores, pasamos a estimar estos modelos.

Finalmente, examinamos los residuales observando la distribución de los errores, su FAC, o algún test que compruebe si los términos de la autocorrelación son diferentes de cero, como el test de Box-Pierce o el de Ljung-Box.

Ej. Llegada de turistas

```r
# Graficamos la serie original
turistas %>% ggplot(aes(x=fecha,y=total)) +
  geom_line(color="red") + theme_gc()

# Observamos la tendencia exponencial y convertimos a log
turistas1 <- turistas %>% mutate(total=log(total)) %>% select(fecha,total)
turistas1 %>% ggplot(aes(x=fecha,y=total)) +
  geom_line(color="red") + theme_gc()

# Diferenciamos
turistas1 <- turistas1 %>%
  mutate(total=c(NA,diff(total))) %>% filter(complete.cases(.))
turistas1 %>%
  ggplot(aes(x=fecha,y=total)) +
  geom_line(color="red") + theme_gc()

# Vemos el FAC y FACP. Se nota muy claramente la estacionalidad.
acf(turistas1$total)
pacf(turistas1$total)
```

Ej. Llegada de turistas

```r
# Probamos entonces varios candidatos

# ARIMA(1,0,0)x(1,0,0)12
arima1 <- arima(turistas1$total, order = c(1,0,0),
     seasonal = list(order=c(1,0,0),period=12))

# ARIMA(0,0,1)x(1,0,0)12
arima2 <- arima(turistas1$total, order = c(0,0,1),
     seasonal = list(order=c(0,0,0),period=12))

# ARIMA(1,0,1)x(1,0,0)12
arima3 <- arima(turistas1$total, order = c(1,0,1),
     seasonal = list(order=c(1,0,0),period=12))

# ARIMA(0,0,0)x(1,0,0)12
arima4 <- arima(turistas1$total, order = c(0,0,0),
     seasonal = list(order=c(1,0,0),period=12))

# Graficar FAC y FACP
par(mfrow=c(2,2))
acf(resid(arima1));pacf(resid(arima1));
acf(resid(arima2));pacf(resid(arima2));
par(mfrow=c(2,2))
acf(resid(arima3));pacf(resid(arima3));
acf(resid(arima4));pacf(resid(arima4));
```

E.j. Llegada de turistas

```r
# Estos errores sugieren que nos falta un término MA estacional

# ARIMA(1,0,1)x(1,0,1)12
arima5 <- arima(turistas1$total, order = c(1,0,1),
     seasonal = list(order=c(1,0,1),period=12))
par(mfrow=c(2,1))
acf(resid(arima5));pacf(resid(arima5))

# test de portmanteau concluye que los errores son independientes
Box.test(resid(arima5), lag=12)

# Sin embargo, errores muestran heteroscedasticidad
plot(resid(arima5))
```

Ejercicios

Analiza las series x_1, x_2, x_3, x_4 y determina los mejores candidatos de modelos.