Análisis de Series Temporales

...

Matías Marenchino - Gabriel Miretti

Bibliografía

- **El material del curso está basado en el libro Forecasting: Principles and Practice (Rob J Hyndman and George Athanasopoulos): https://otexts.org/fpp2/**

- **La mayoría de las slides son de los mismos autores: https://github.com/robjhyndman/ETC3550Slides**

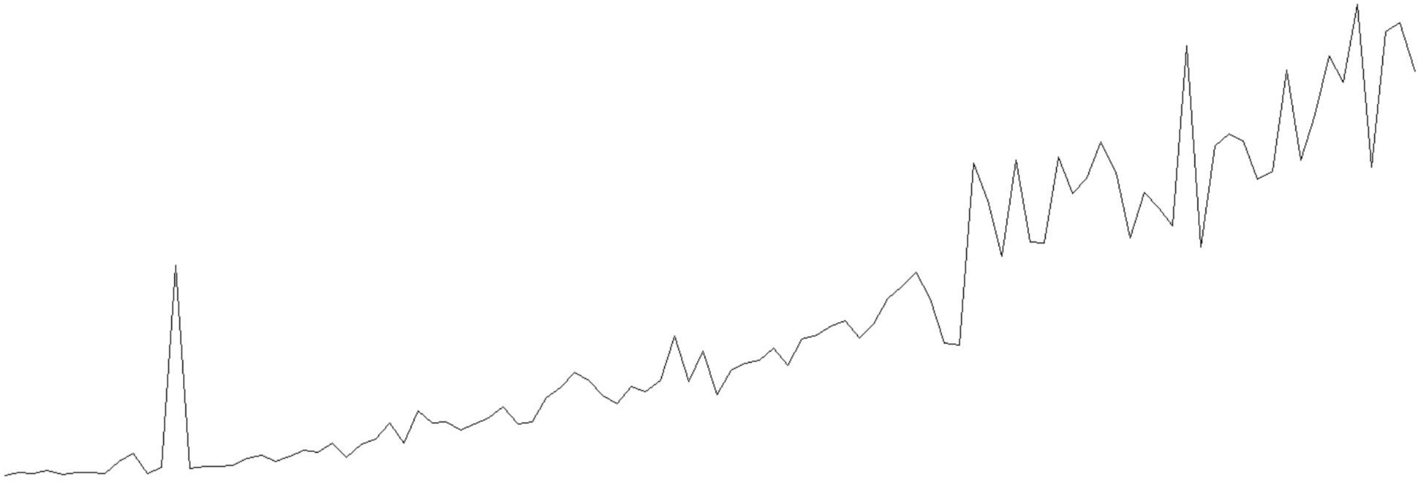
- **Nos gusta python (y no nos gusta R), así que usaremos el material que está acá: https://github.com/gmiretti/forecasting, que está basado en https://github.com/mscharth/forecasting**

RoadMap

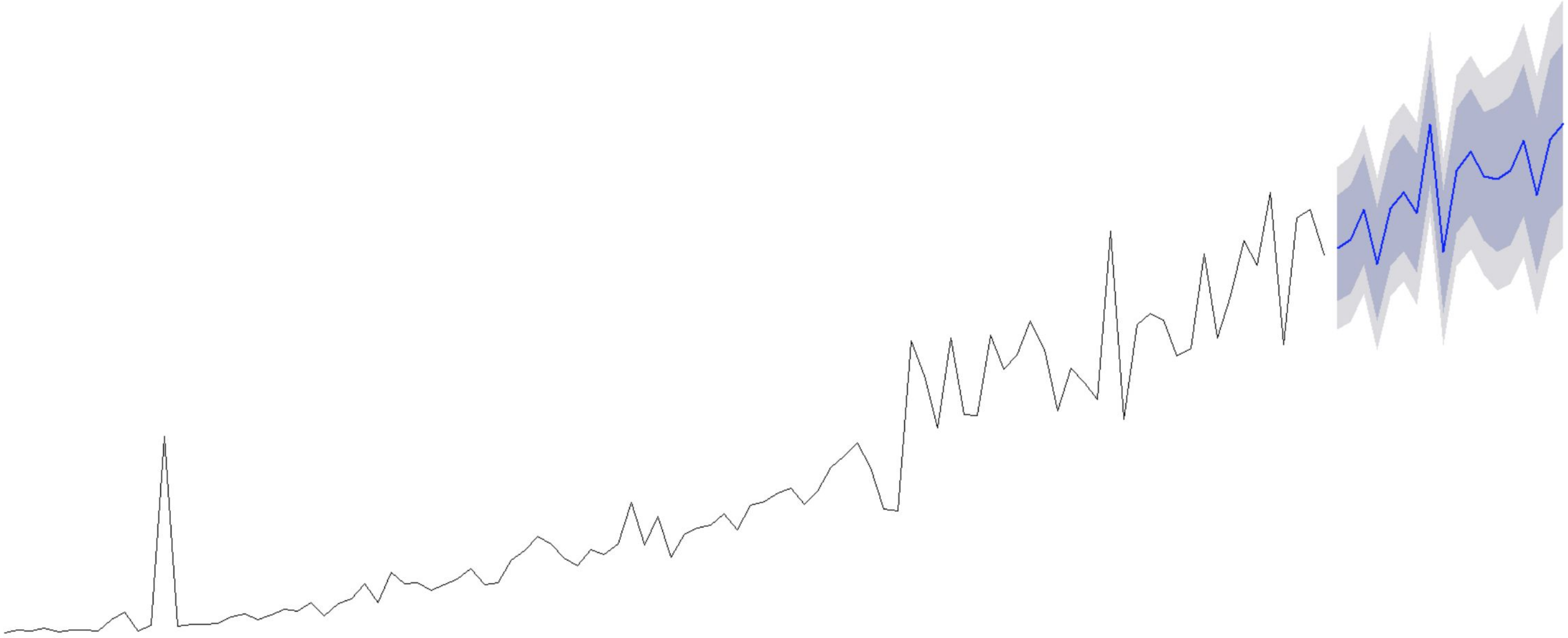
- **Introducción**
- **Definiciones / Gráficos / Patrones**
- **Análisis Predictivo: Métodos Básicos**
- **Evaluación de modelos**
- **Regresión**
- **Holt Winters**
- **ARIMA**
- **Bonus track: Popurrí de cosas "interesantes" (Multivariado/Anomalías/LSTM/Prophet)**

Introducción

El problema



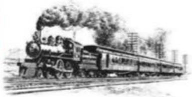
El problema



Por qué no usar lo que sabemos de las otras materias de diplodatos?

Predecir/Pronosticar (Forecast) puede ser difícil

1800



"Rail travel at high speed is not possible, because passengers, unable to breathe, would die of asphyxia."

Dr. Dionysys Larder, Professor of Natural Philosophy & Astronomy, University College London

1859

"Drill for oil? You mean drill into the ground to try and find oil? You're crazy!"

Associates of Edwin L. Drake refusing his suggestion to drill for oil in 1859 (Later that year, Drake succeeded in drilling the first oil well.)

1876

"This telephone has too many shortcomings to be seriously considered as a means of communication."

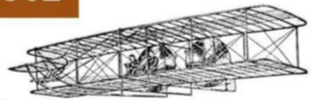
Western Union internal memo

1880

"Everyone acquainted with the subject will recognize it as a conspicuous failure."

Henry Morton, president of the Stevens Institute of Technology, on Edison's light bulb

1902



"Flight by machines heavier than air is unpractical and insignificant, if not utterly impossible."

Simon Newcomb, Canadian-American astronomer and mathematician, 18 months before the Wright Brothers' flight at Kittyhawk

1903

"The horse is here to stay, but the automobile is only a novelty, a fad."

The president of the Michigan Savings Bank, advising Henry Ford's lawyer not to invest in the Ford Motor Company

1916

"The idea that cavalry will be replaced by these iron coaches is absurd. It is little short of treasonous."

Comment of Aide-de-camp to Field Marshal Haig, at tank demonstration

1916

"The cinema is little more than a fad. It's canned drama. What audiences really want to see is flesh and blood on the stage."

Charlie Chaplin, actor, producer, director, and studio founder

1921

"The wireless music box has no imaginable commercial value. Who would pay for a message sent to no one in particular?"

Associates of commercial radio and television pioneer, David Sarnoff, responding to his call for investment in the radio

1946

"Television won't last because people will soon get tired of staring at a plywood box every night."

Darryl Zanuck, movie producer, 20th Century Fox

1977

"There is no reason for any individual to have a computer in his home."

Ken Olson, president, chairman and founder of Digital Equipment Corporation

1995

"The truth is no online database will replace your daily newspaper..."

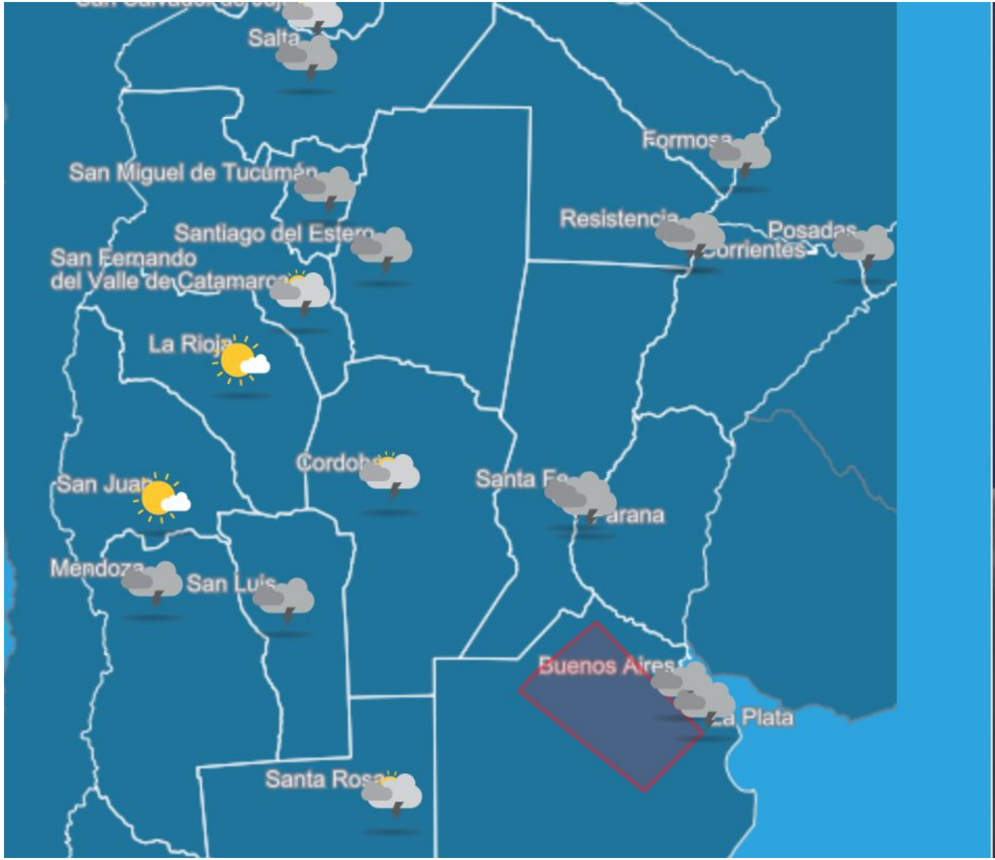
Clifford Stoll, Newsweek article entitled ***The Internet? Bah!***

Qué podemos pronosticar?

Qué podemos pronosticar?

Qué podemos pronosticar?

Qué podemos pronosticar?



Qué es más fácil de pronosticar?

- Demanda energética en los próximos 3 días?
- Qué números van a salir en el próximo sorteo del Quini6?
- El precio de las acciones de Google en 6 meses?
- Temperatura máxima de mañana?
- La hora de salida del sol en exactamente un año?
- El precio del dólar la semana que viene?

Qué es más fácil de pronosticar?

- Demanda energética en los próximos 3 días?
- Qué números van a salir en el próximo sorteo del Quini6?
- El precio de las accior
- Temperatura máxima (
- La hora de salida del sol en exactamente un año?
- El precio del dólar la semana que viene?

- Cómo medimos esa "facilidad"?
- Cuándo algo es fácil o difícil de pronosticar?

Qué es más fácil de pronosticar?

Algo es "fácil" de pronosticar si:

- Entendemos los factores que contribuyen al fenómeno
- Tenemos datos disponibles para analizar
- Los mismos pronósticos no afectan el fenómeno a predecir
- El fenómeno tiene poca aleatoriedad
- El comportamiento en el futuro es similar al comportamiento en el pasado

Series Temporales

Definición

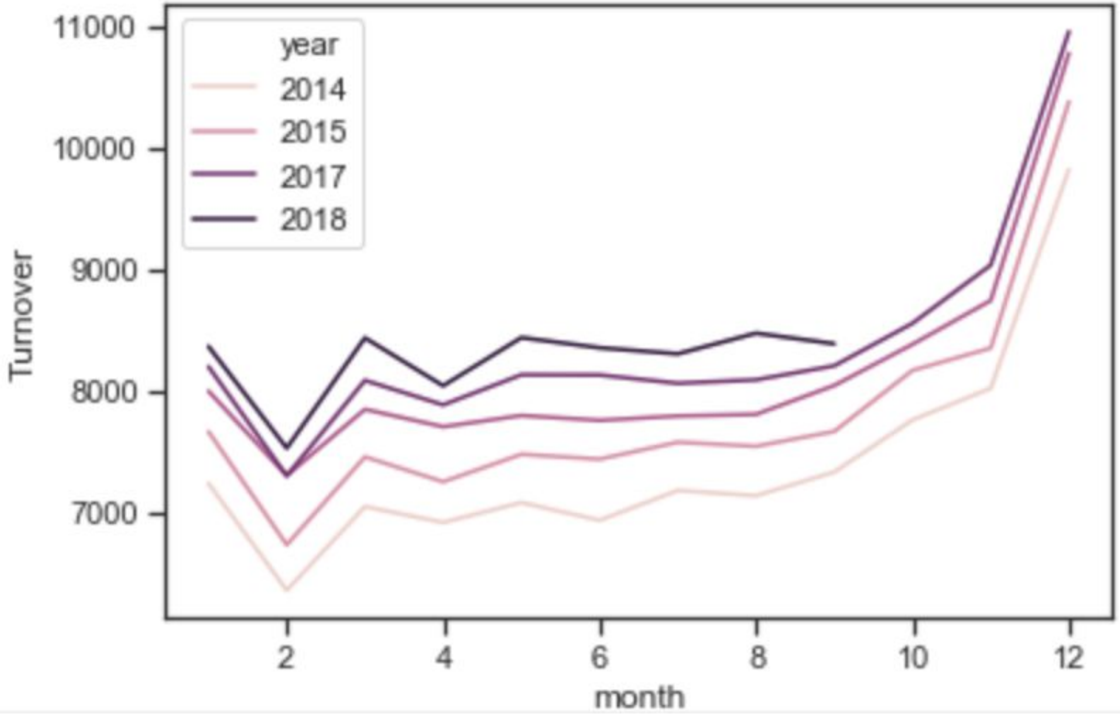
Una serie temporal es un conjunto de observaciones

$$y_1 , y_2 , \cdots , y_t$$

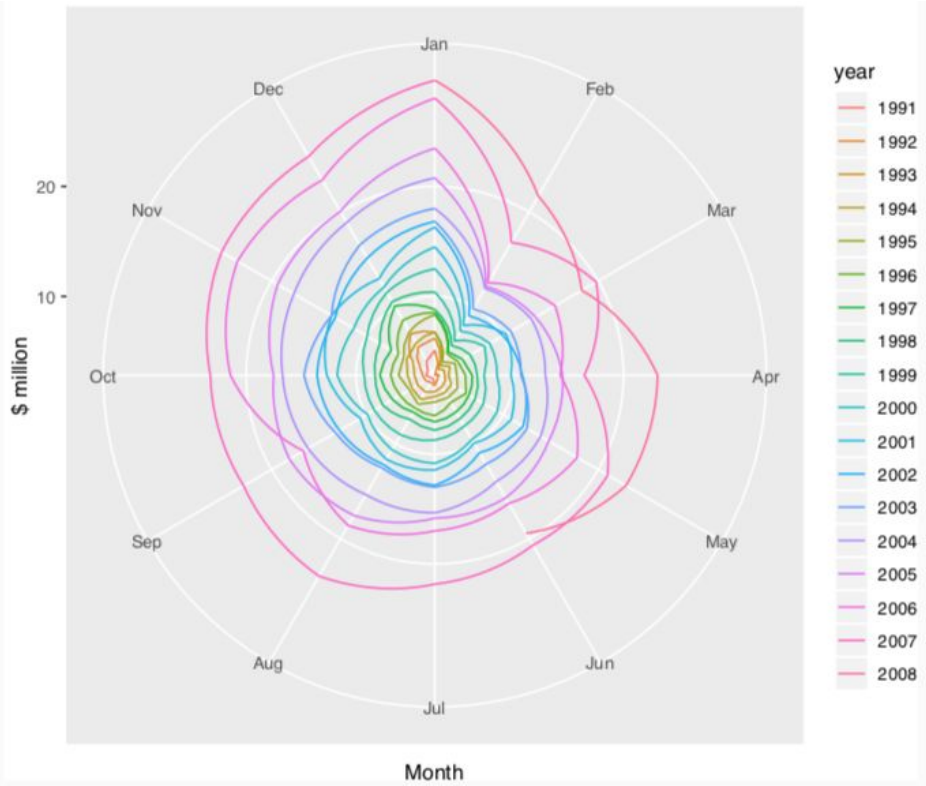
que están ordenadas en el tiempo. Usualmente vienen acompañadas por una etiqueta de tiempo.

Cómo las visualizamos?
Time plots

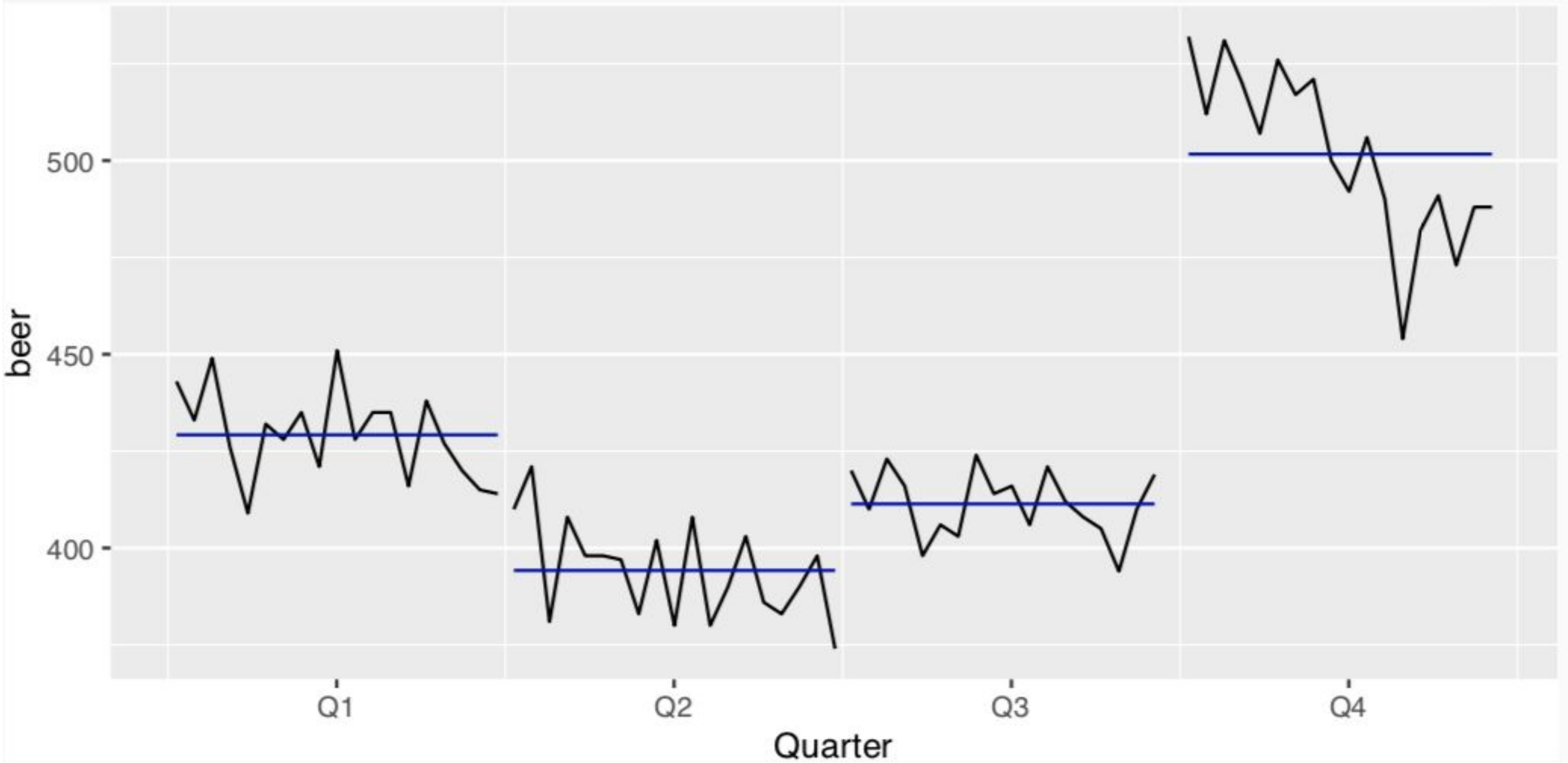
Cómo las visualizamos?
Seasonal plots



Cómo las visualizamos?
Seasonal polar plots

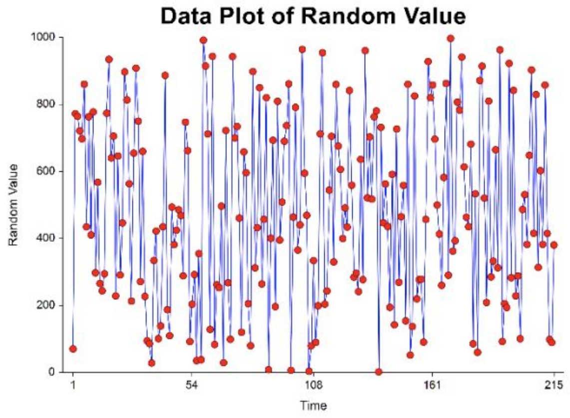


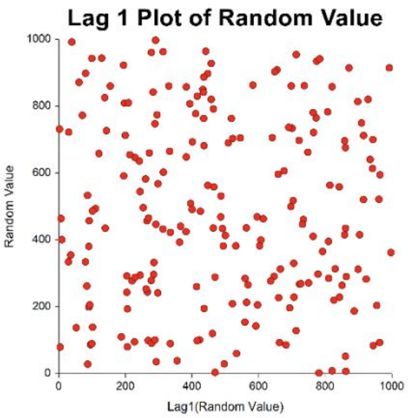
Cómo las visualizamos?
Seasonal subseries plots



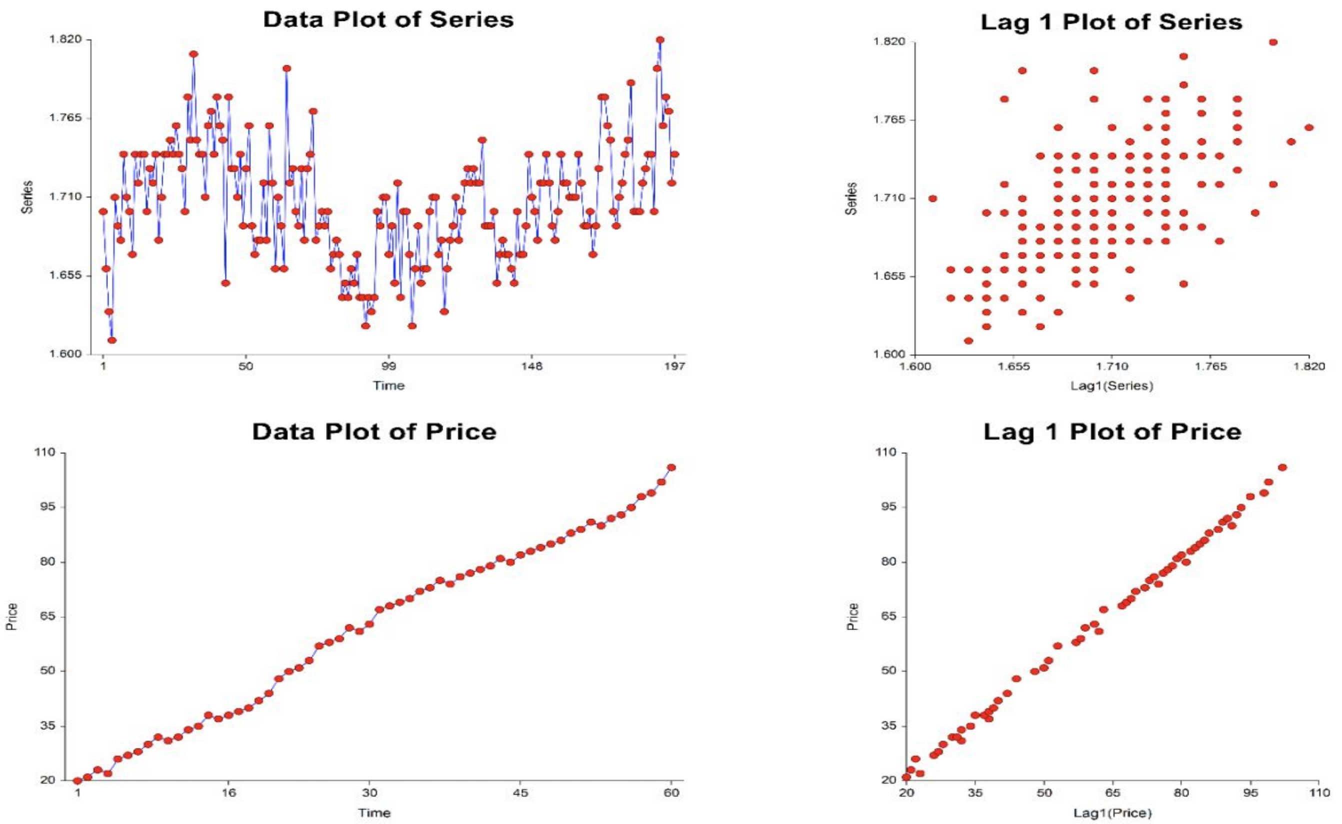
Cómo las visualizamos?
Lag plots

$\mathrm{Lag}_k(y_t) = y_{t-k}$ entonces graficamos y_t vs y_{t-k} para distintos valores de k.





Cómo las visualizamos? Lag plots



Predicción (Forecast)

Nuestro objetivo es el de predecir el valor de una variable en instante futuro $\{t + h\}$, dada una series temporal con observaciones hasta el instante t. Es decir, queremos predecir,

$$y_{t+h}$$

dados

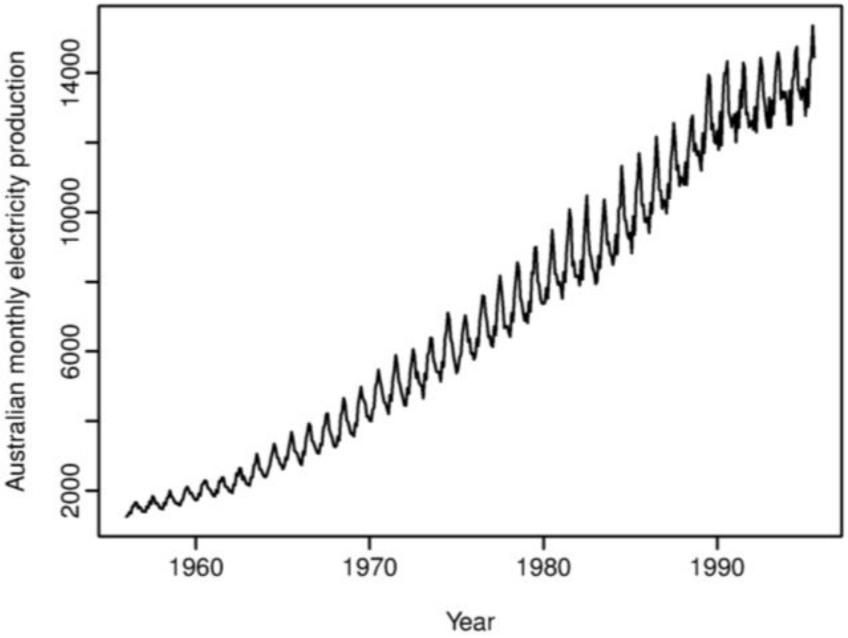
$$y_1, y_2, \cdots, y_t \quad \text{o} \quad y_{1:t}$$

h se suele llamar "horizon". Además, usualmente a las predicciones las denotaremos como

$$\hat{y}_{t+h}$$

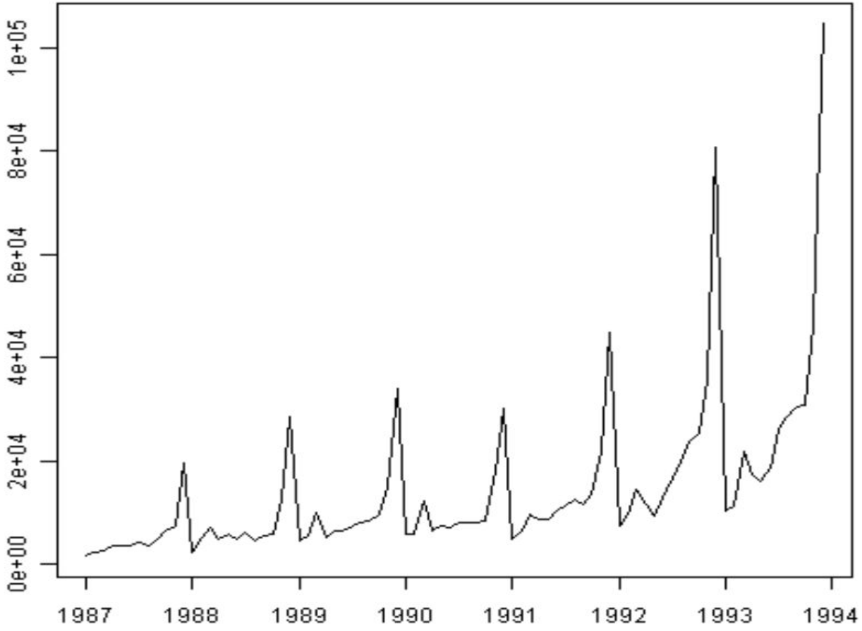
Patrones

- **Tendencia:** incremento o decremento sistemático a largo plazo (no necesariamente lineal)



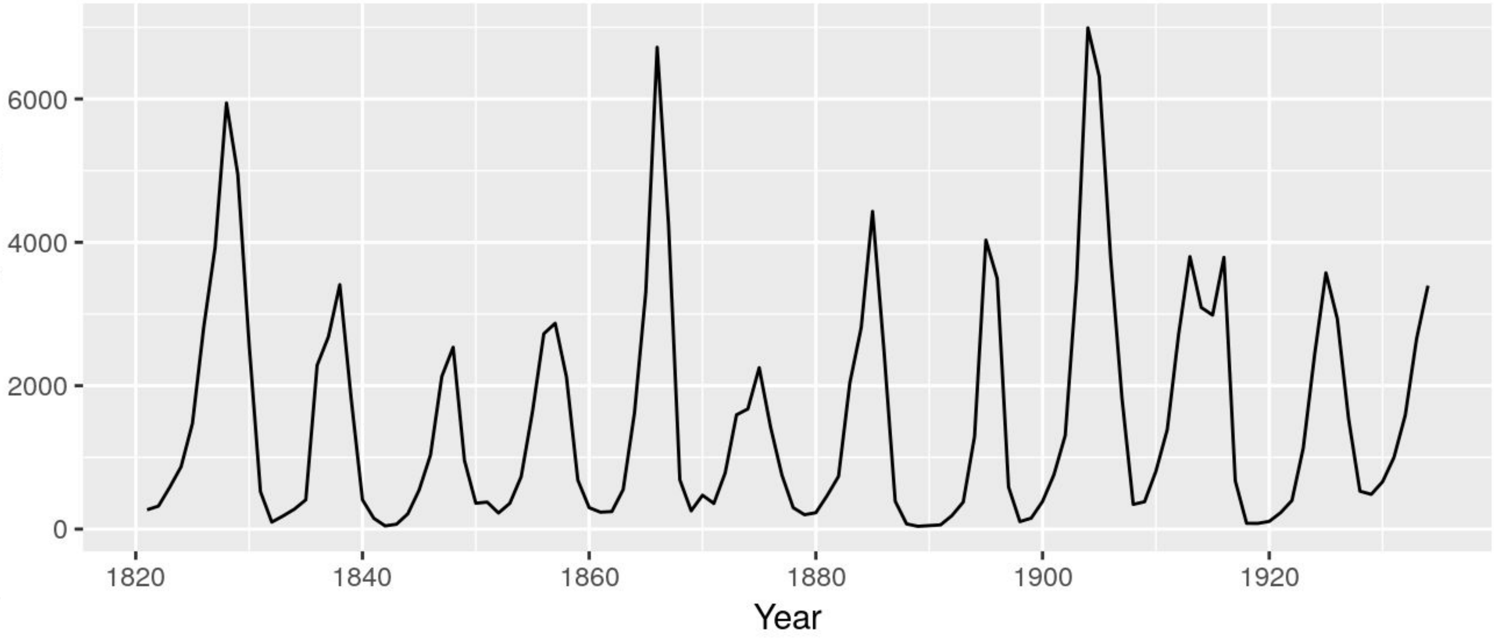
Patrones

- **Estacional:** el patrón ocurre cuando la serie está afectada por razones estacionales (ciertos períodos de tiempo)



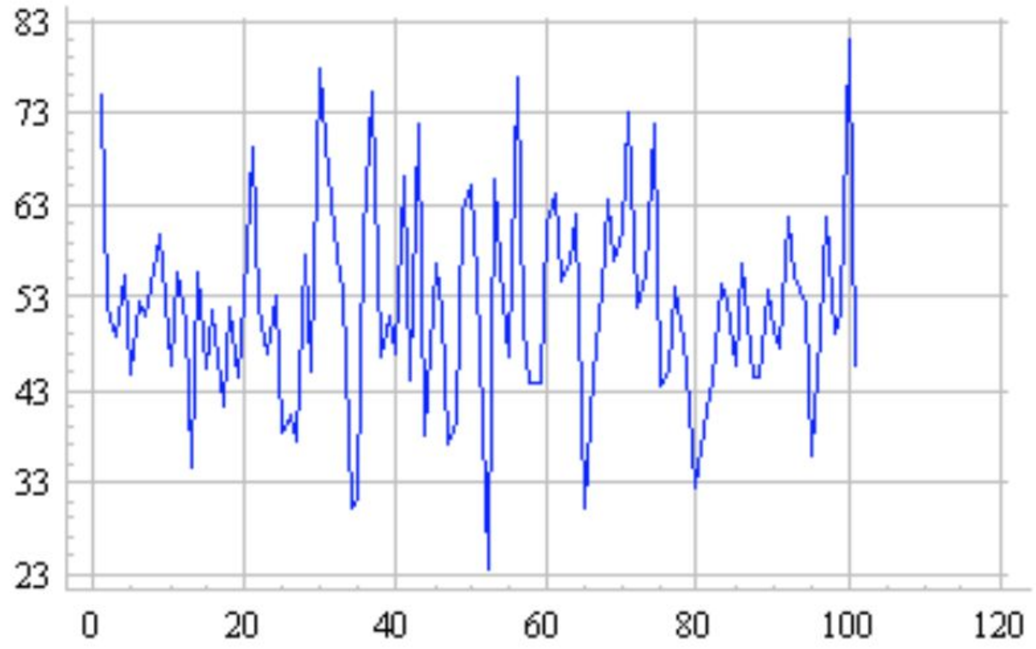
Patrones

- **Cíclico:** se da cuando hay fluctuaciones de mediano o largo plazo y que no tienen una frecuencia fija (o estable)

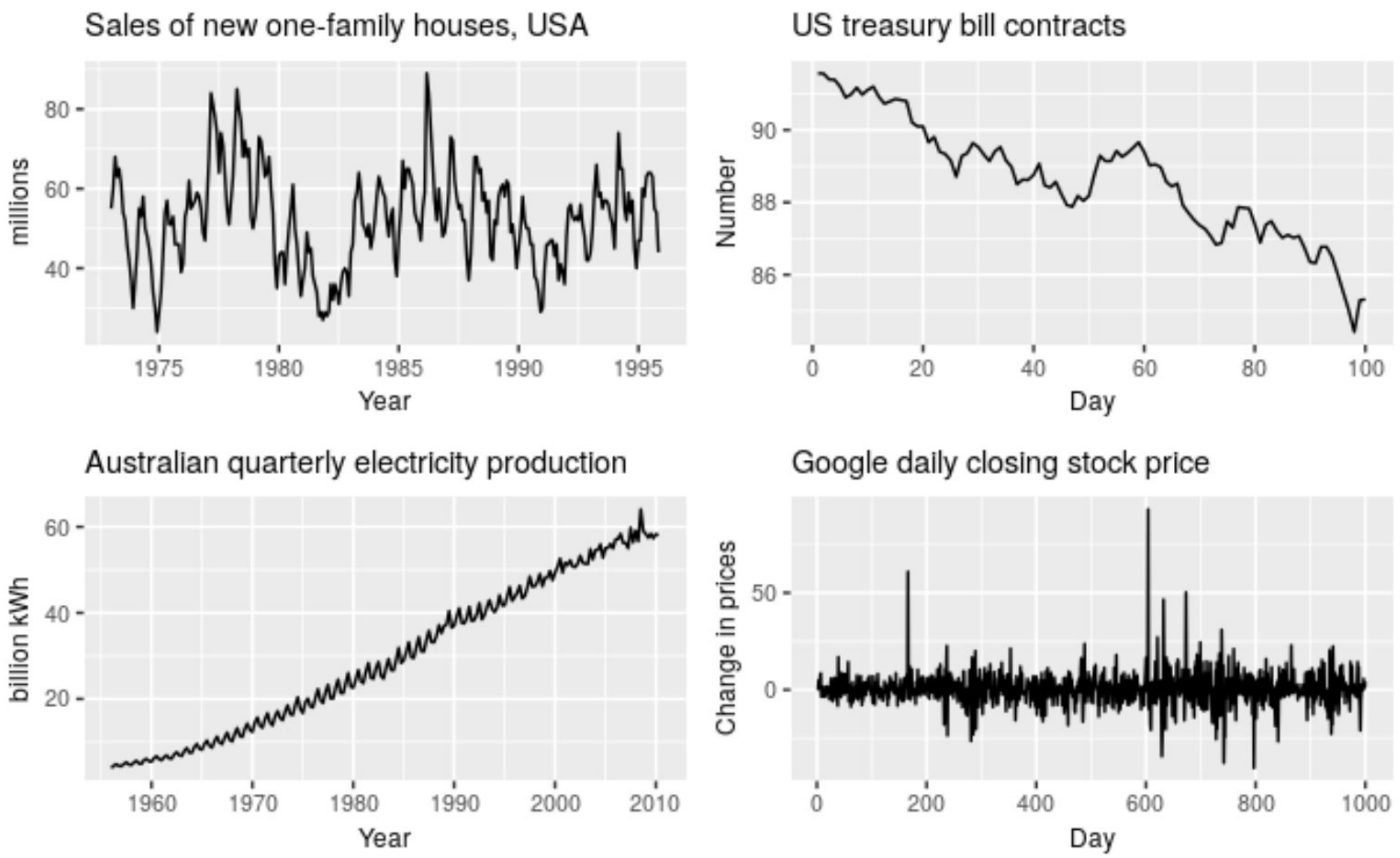


Patrones

- **Irregular:** hay fluctuaciones de corto plazo y mucho "ruido"



Patrones



Sales of new one-family houses, USA
millions
80
60
40
1975
1980
1985
1990
1995
Year
US treasury bill contracts
Number
90
88
86
0
20
40
60
80
100
Day
Australian quarterly electricity production
billion kWh
60
40
20
1960
1970
1980
1990
2000
2010
Year
Google daily closing stock price
Change in prices
50
0
0
200
400
600
800
1000
Day

Modelos de Series Temporales

- **Aditivo:**

$$y_t = T_t + S_t + C_t + E_t$$

- **Multiplicativo:**

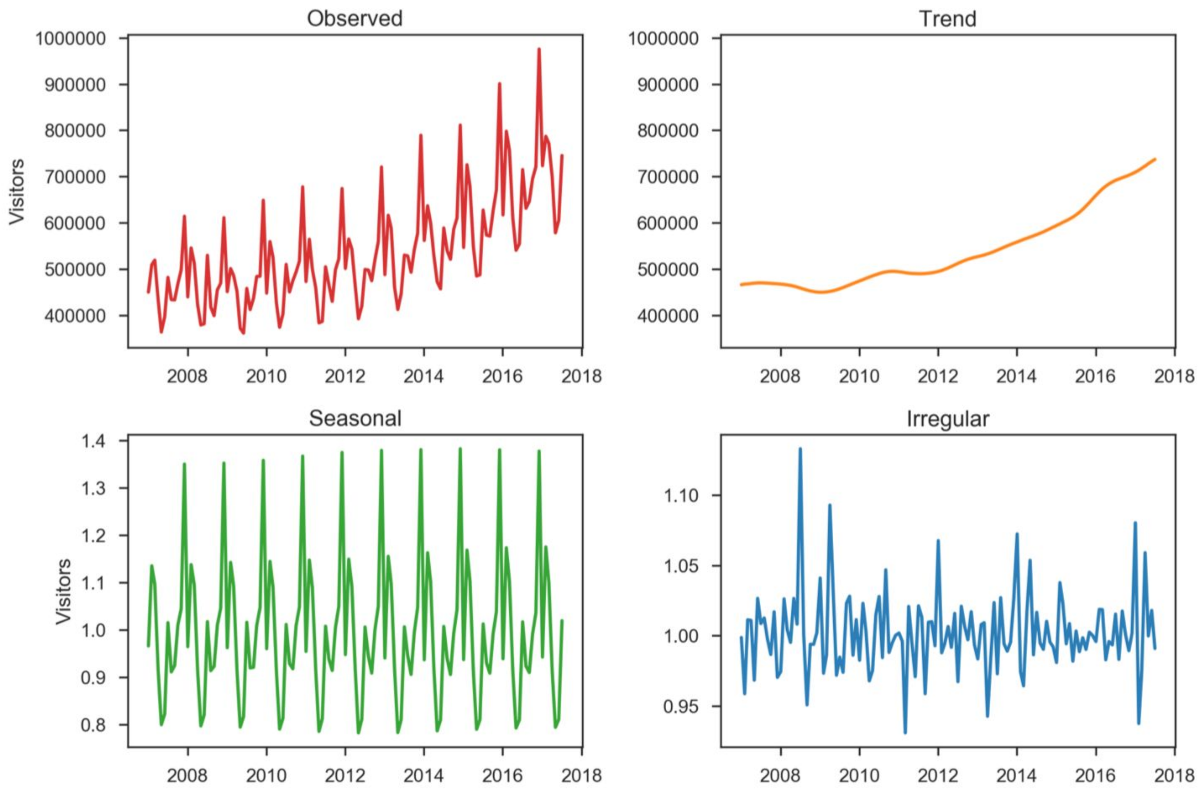
$$y_t = T_t . S_t . C_t . E_t$$

Métodos de descomposición

- Los métodos de descomposición para series temporales son algoritmos para dividir las series en distintas componentes. Usualmente, esta aproximación se usa para ajustes estacionales.

- Estos métodos suelen ser útiles para EDA (análisis exploratorio de datos). Permiten visualizar patrones en los datos.

Métodos de descomposición

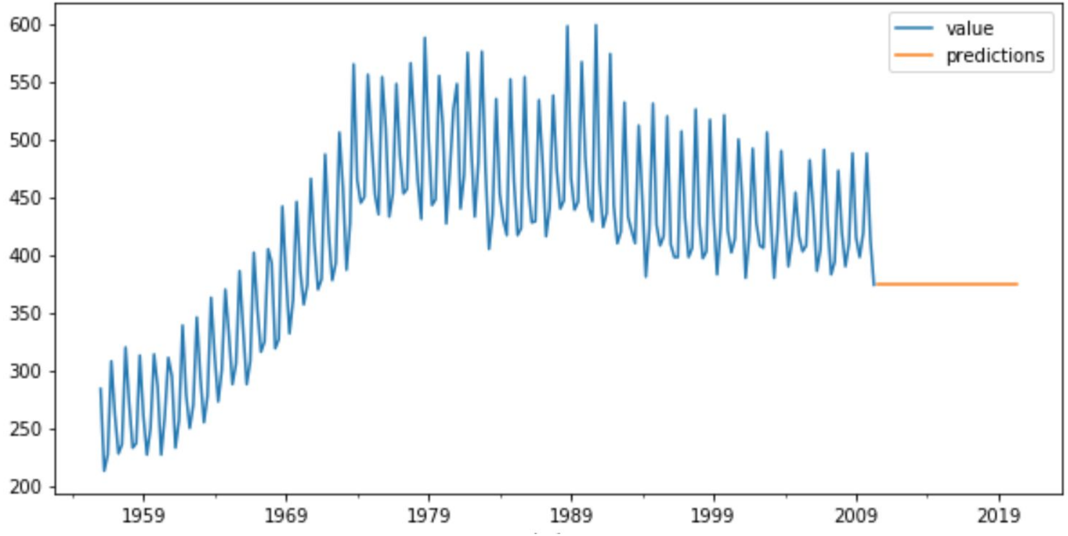


Análisis Predictivo
Métodos Básicos

Naïve (Random walk)

Para predecir un valor, se usa la última observación de la serie temporal:

$$\hat{y}_{t+h} = y_t$$



Seasonal Naïve (Seasonal Random walk)

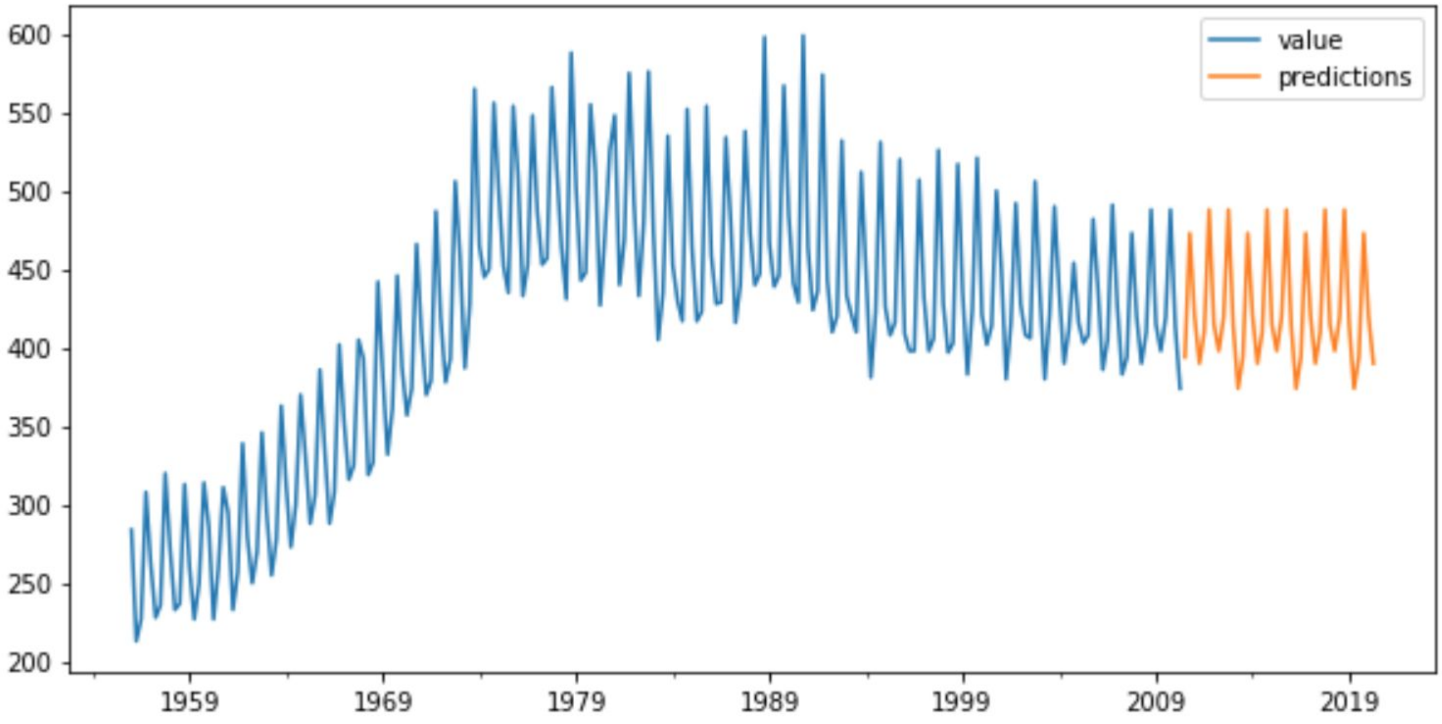
Si la serie presenta comportamientos estacionales, en lugar de usar la última observación, usamos el último valor correspondiente a la misma "estación" actual:

$$\hat{y}_{t+h} = y_{t+h-km}$$

Donde *m* representa la longitud de la estación y *k* es el "salto" a la estación correspondiente:

$$k = \left\lfloor \frac{h-1}{m+1} \right\rfloor$$

Seasonal Naïve (Seasonal Random walk)

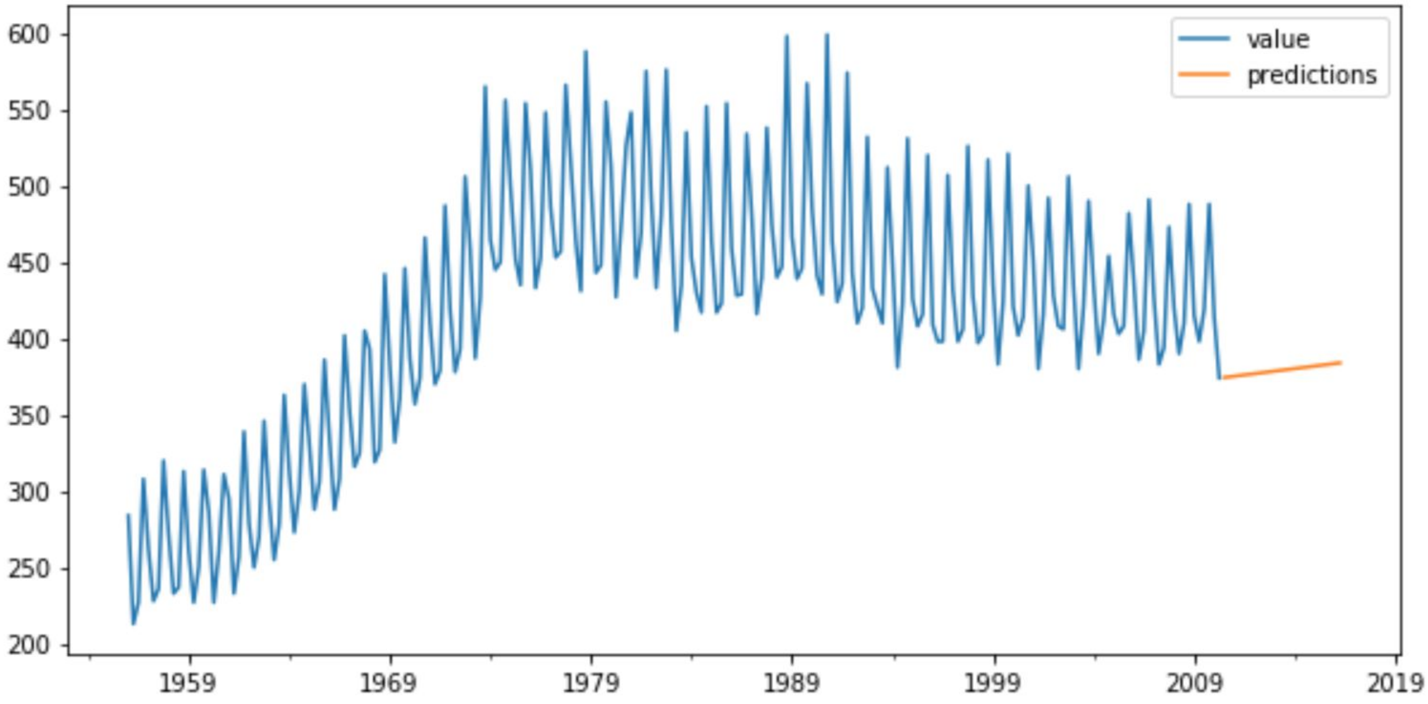


Drift

Adaptación del método Naïve, que permite que la predicción crezca o decrezca dependiendo del promedio de la serie y de cuánto nos alejamos en el tiempo de la última observación

$$\hat{y}_{t+h} = y_t + h \sum_{i=2}^{t} \frac{y_i - y_{i-1}}{t-1}$$

Drift



value
predictions
600
550
500
450
400
350
300
250
200
1959
1969
1979
1989
1999
2009
2019

Evaluación de Modelos

...

Autocorrelación

La ***correlación*** entre dos variables mide la relación lineal que existe entre éstas.

La **autocorrelación** mide la relación lineal que existe entre una serie temporal $\boldsymbol{y_t}$ y sus valores retrasados (lagged). Es decir, compara

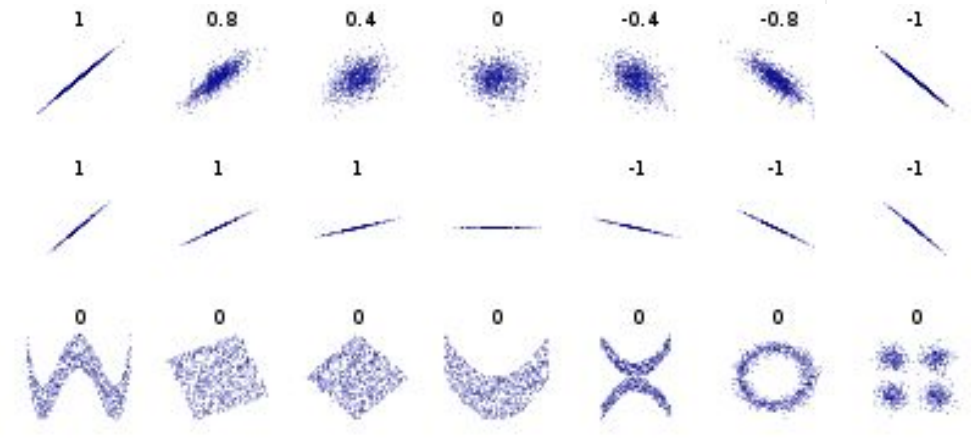
$$y_t \text{ vs } y_{t+k}$$

Para hacerlo:

$$\rho_k = \frac{E[(Y_t-\mu)(Y_{t-k}-\mu)]}{\sigma^2} = \mathrm{Corr}(Y_t, Y_{t+k})$$

Autocorrelación

La ***correlación*** entre dos variabl
entre éstas.



La **autocorrelación** mide la rela
temporal $\boldsymbol{y_t}$ y sus valores retrasa

$$y_t \text{ VS } y_{t+k}$$

Para hacerlo:

$$\rho_k = \frac{E[(Y_t - \mu)(Y_{t-k} - \mu)]}{\sigma^2} = \mathrm{Corr}(Y_t, Y_{t+k})$$

Autocorrelación

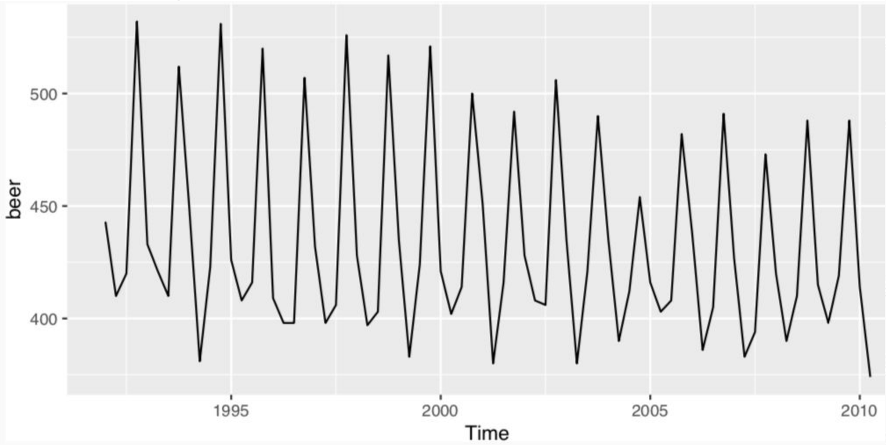
La autocorrelación muestral se define como

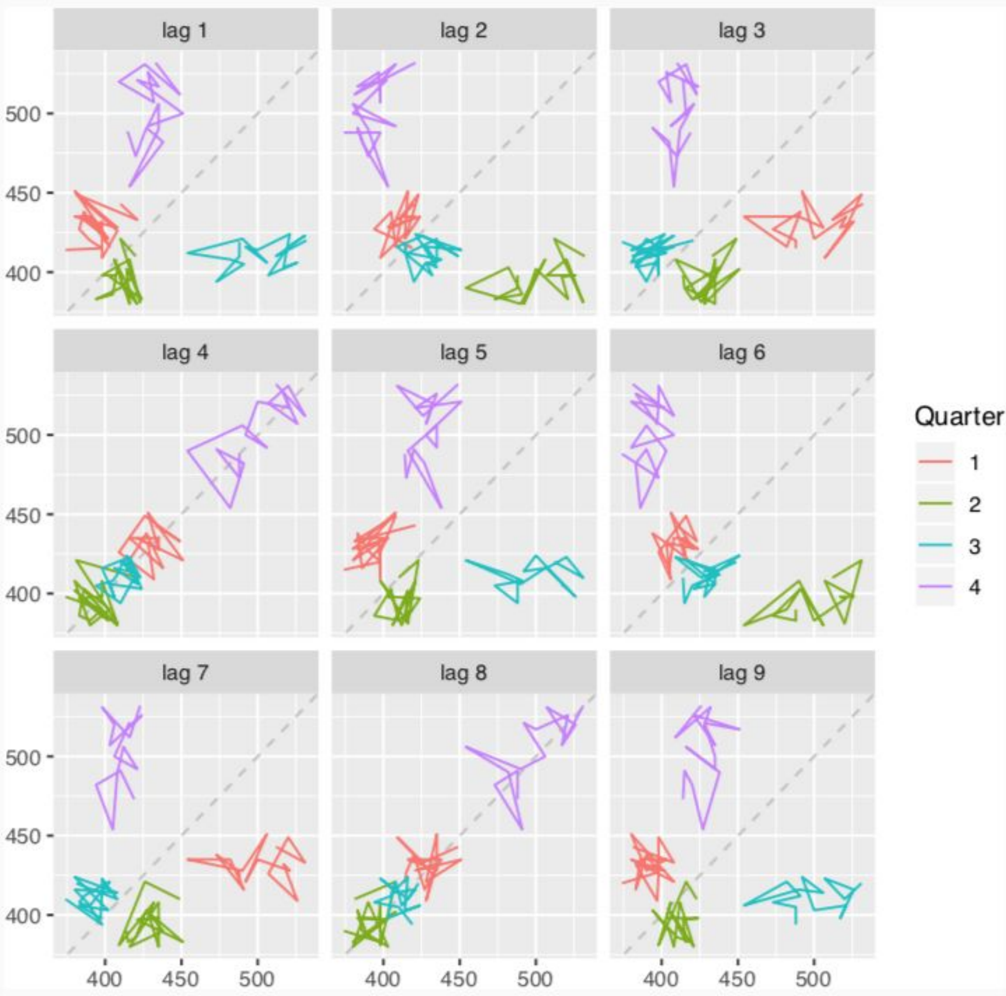
$$r_k = \frac{\sum_{t=1}^{T-k}(y_{t+k}-\bar{y})(y_t-\bar{y})}{\sum_{t=1}^{T}(y-\bar{y})^2}$$

Así,

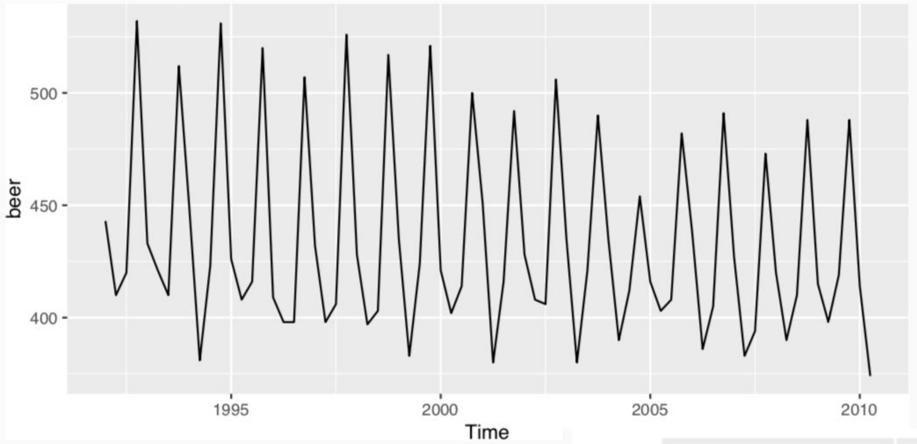
- r_1 **indica cómo se relacionan valores consecutivos en la serie**
- r_2 **indica cómo se relacionan valores separados por 2 instantes de tiempo**
- **...**

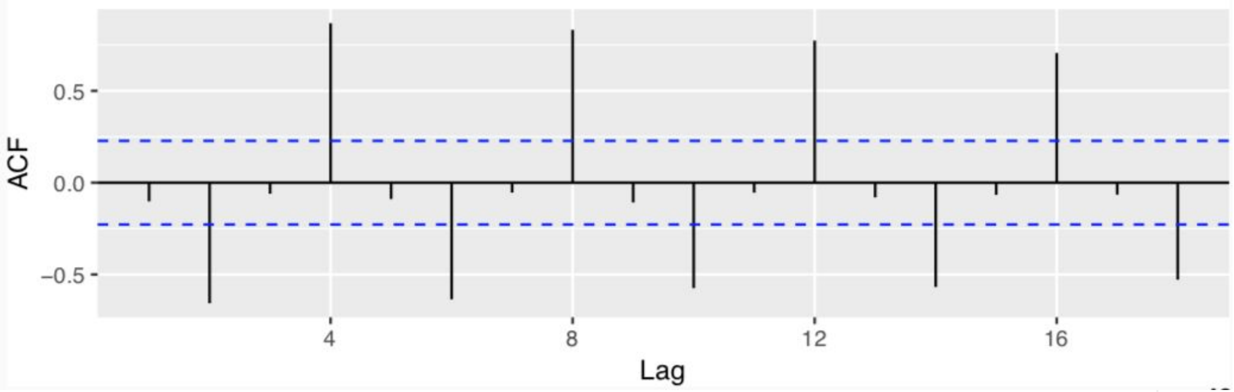
Autocorrelación



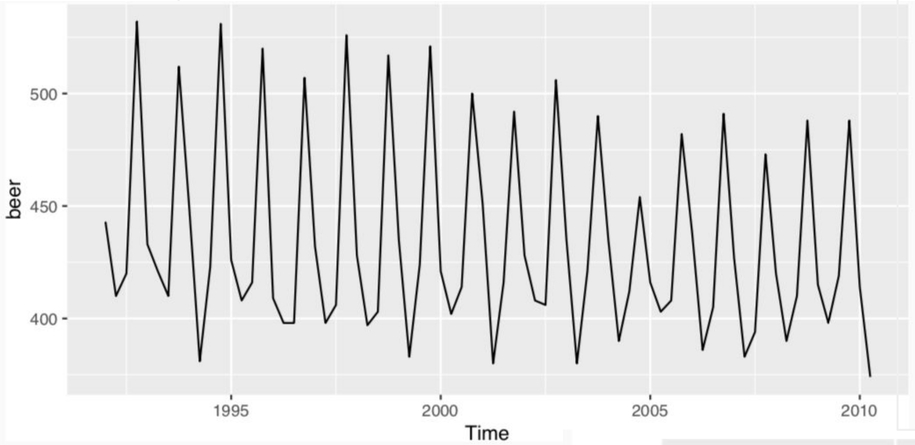


Autocorrelación



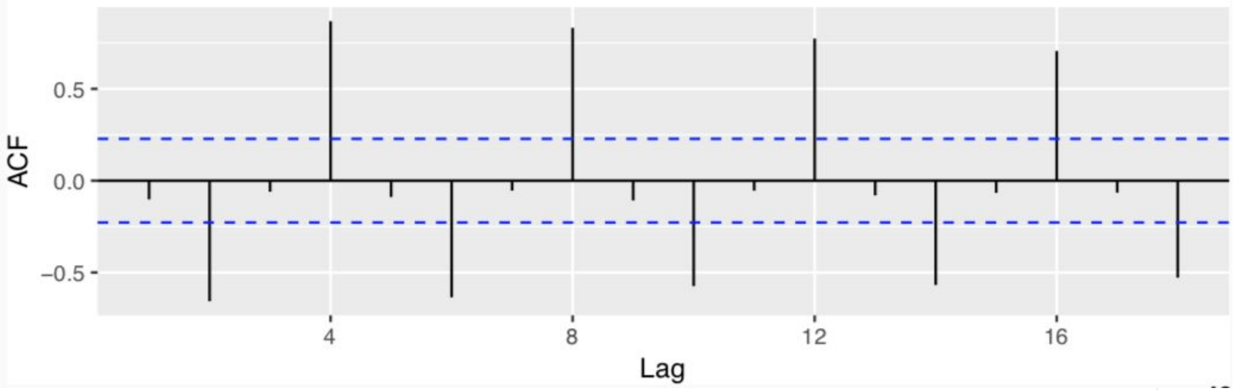


Autocorrelación

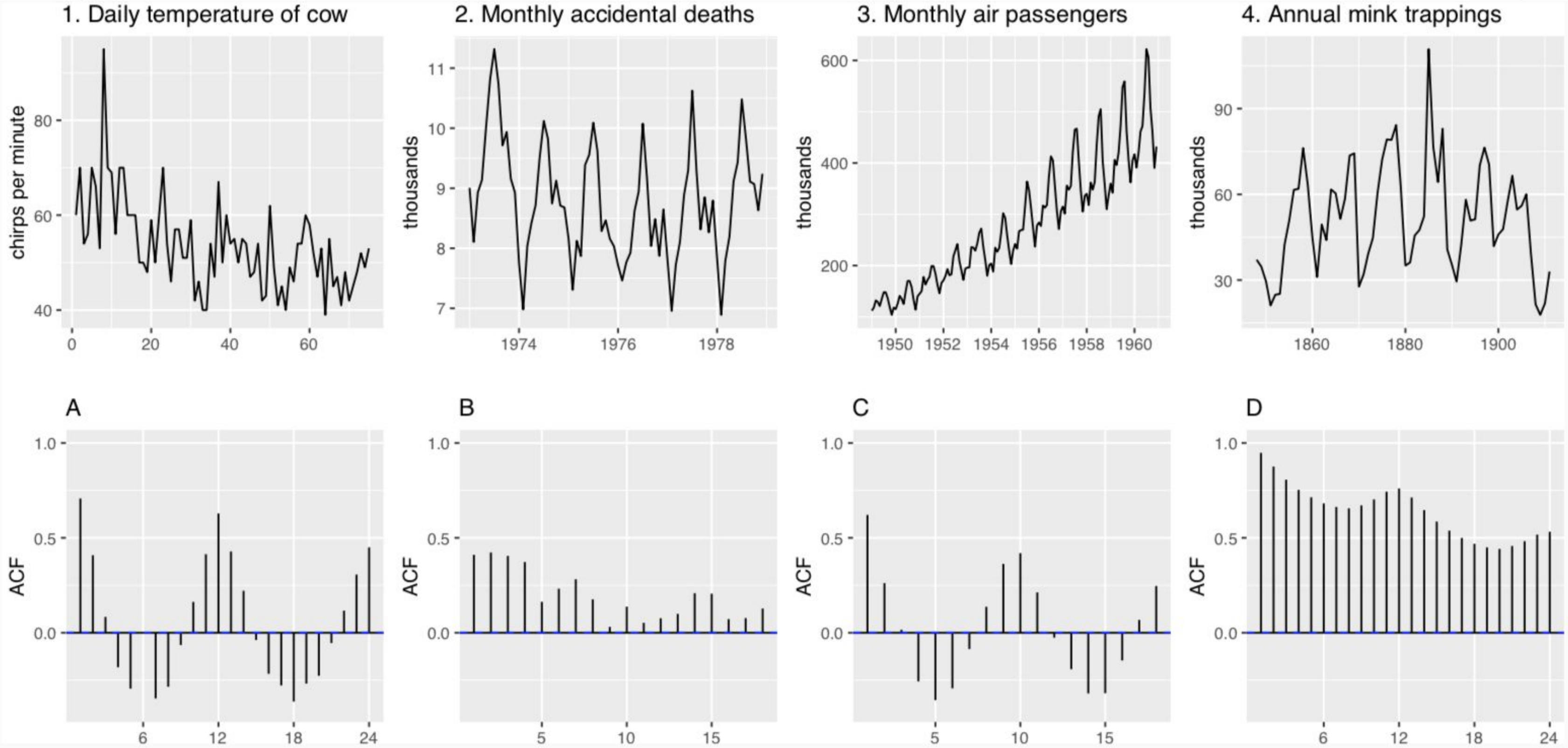


- **Tendencia:** cuando ocurre el fenómeno, las autocorrelaciones para retrasos (lags) pequeños tienden a ser grandes y positivas

- **Estacional:** en este caso las autocorrelaciones son grandes en los retrasos correspondientes a esas estaciones



Autocorrelación: cuál es cuál?



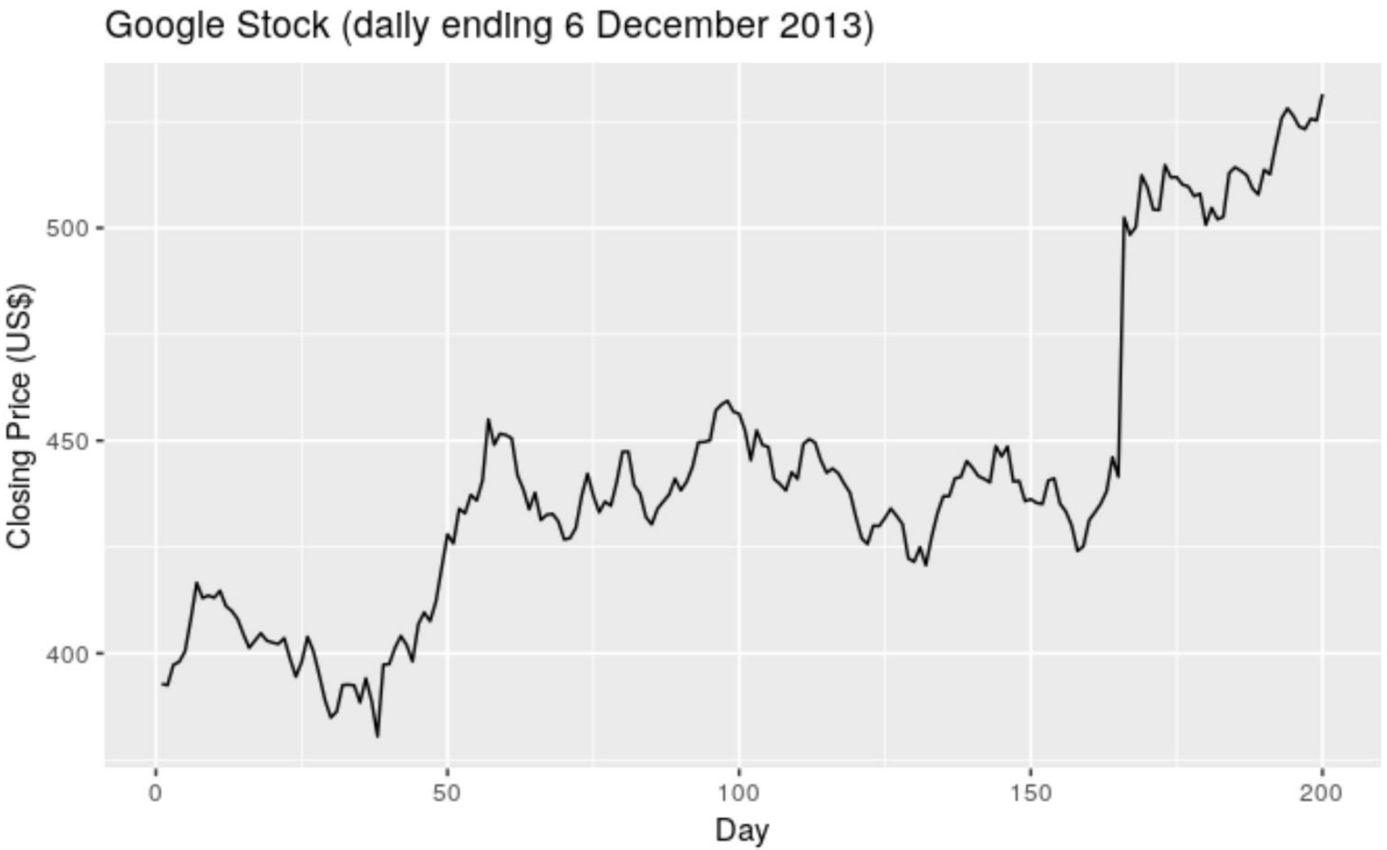
Análisis de Residuos

Los valores residuos corresponden a la diferencia entre los valores observados y los valores predichos:

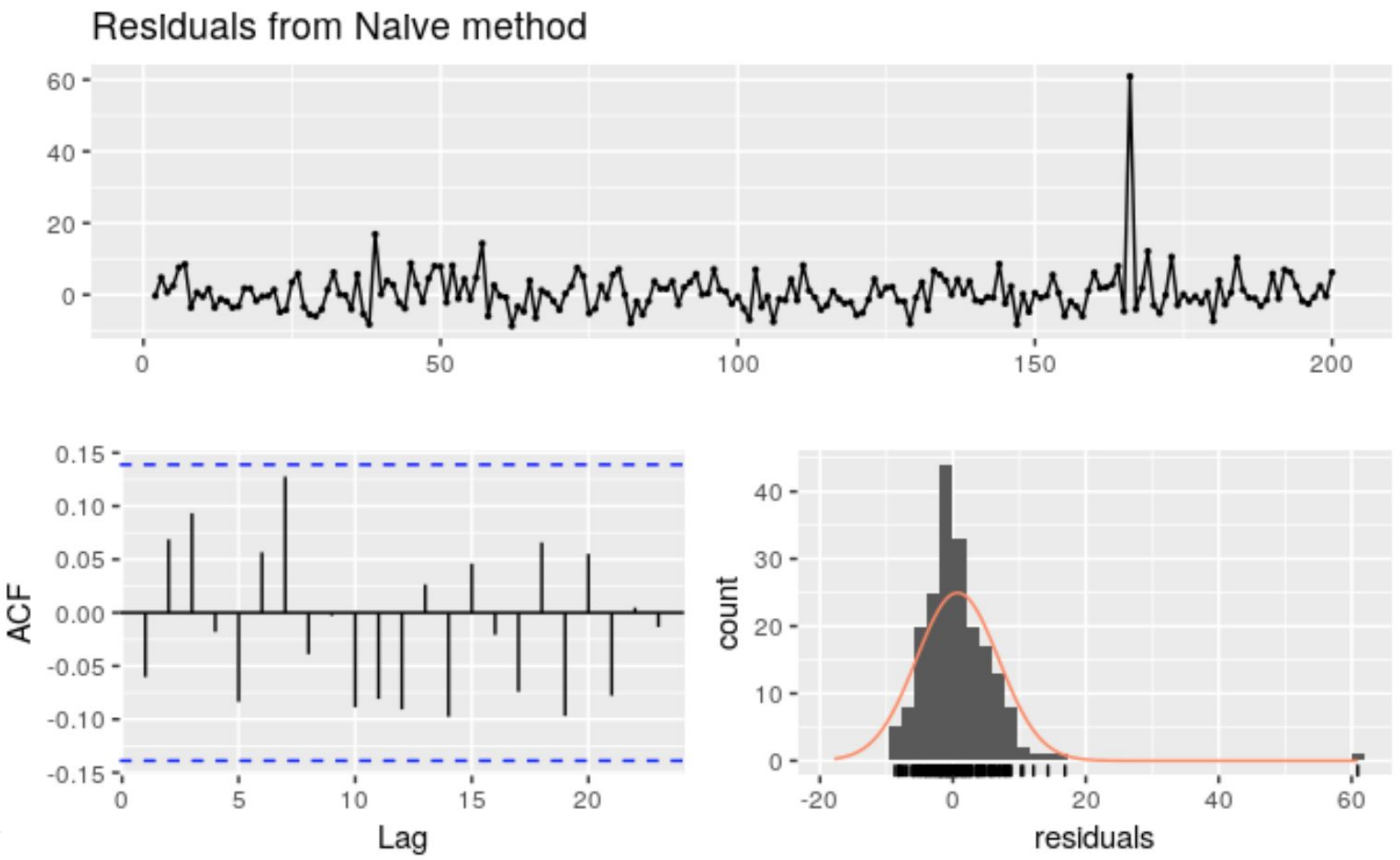
$$e_t = y_t - \hat{y}_t$$

- **Gráfico de residuos:** la presencia de patrones en el time plot de los residuos (como serie temporal) sugiere una predicción deficiente. $\{e_t\}$ debe tener media 0 y valores no correlacionados
- **Gráfico de autocorrelación para residuos:** si el modelo está bien especificado, las autocorrelaciones deberían ser pequeñas, similares a ruido
- **Gráfico de la distribución de residuos:** a priori, nos interesa que los residuos estén normalmente distribuidos (ya veremos el porqué)

Análisis de Residuos



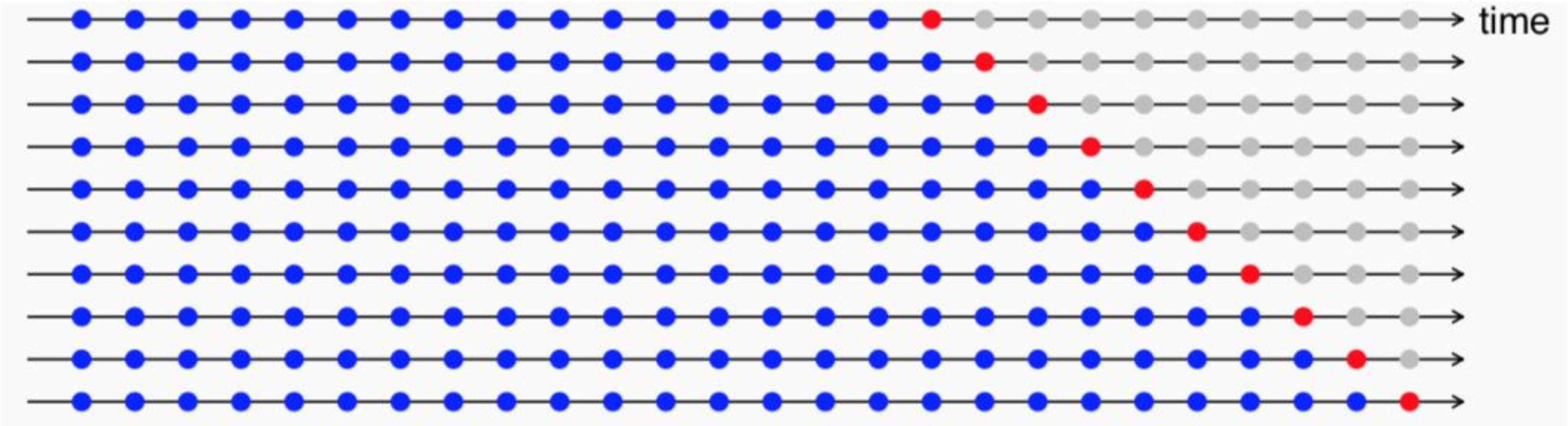
Análisis de Residuos



Validación

- Cuestiones *"obvias"*:
  - Los datos de test no se pueden usar para ajustar las predicciones
  - La evaluación de los modelos se debe hacer únicamente sobre los datos de test
- Habitualmente se asigna entre un 20% y un 50% de los datos al test set.
- Se suele llamar como "in-sample" al training set y "out-sample" al testing set

Validación Real Time (Cross-Validation)



- **Expanding window:** en cada paso, se extiende el training set
- **Rolling window:** en cada paso, se usan solo las últimas n observaciones (los datos del pasado -lejano- pierden relevancia y por eso se ignoran).

Cómo medimos la precisión?

Error:

$$e_{t+h} = y_{t+h} - \hat{y}_{t+h|t}$$

donde el training set está dado por $\{y_i\}_{i<t}$.

El error difiere de los residuos:

- Los residuos se calculan en el training set, los errores en el test
- Los residuos se basan en predicciones de un paso, mientras que los errores pueden incluir múltiples pasos.

Mean Absolute Error (MAE): $\mathrm{mean}(|e_t|)$

Root Mean Squared Error(RMSE): $\sqrt{\mathrm{mean}(e_t^2)}$

Cómo medimos la precisión?

Error porcentual:

$$p_t = 100\,\frac{y_t - \hat{y}_t}{y_t}$$

Mean Absolute Porcentual Error (MAPE): $\mathrm{mean}(|p_t|)$

Ventaja:

Desventaja:

Cómo medimos la precisión?

Error porcentual:

$$p_t = 100\frac{y_t - \hat{y}_t}{y_t}$$

Mean Absolute Porcentual Error (MAPE): $\text{mean}(|p_t|)$

Ventaja: no depende de la escala

Desventaja: problemas con los ceros

Cómo medimos la precisión?

Error escalados (para series no estacionales):

$$q_t = \frac{y_t - \hat{y}_t}{\frac{1}{T-1}\sum_{i=2}^{T} |y_i - y_{i-1}|}$$

como el numerador y el denominador tienen la misma escala, q_t pierde la escala. Un error escalado es menor que 1 si es mejor que el predictor que asigna el promedio.

Error escalados (para series estacionales):

$$q_t = \frac{y_t - \hat{y}_t}{\frac{1}{T-m}\sum_{i=m+1}^{T} |y_i - y_{i-m}|}$$

MASE: $\text{mean}(|q_t|)$

Intervalos de Predicción

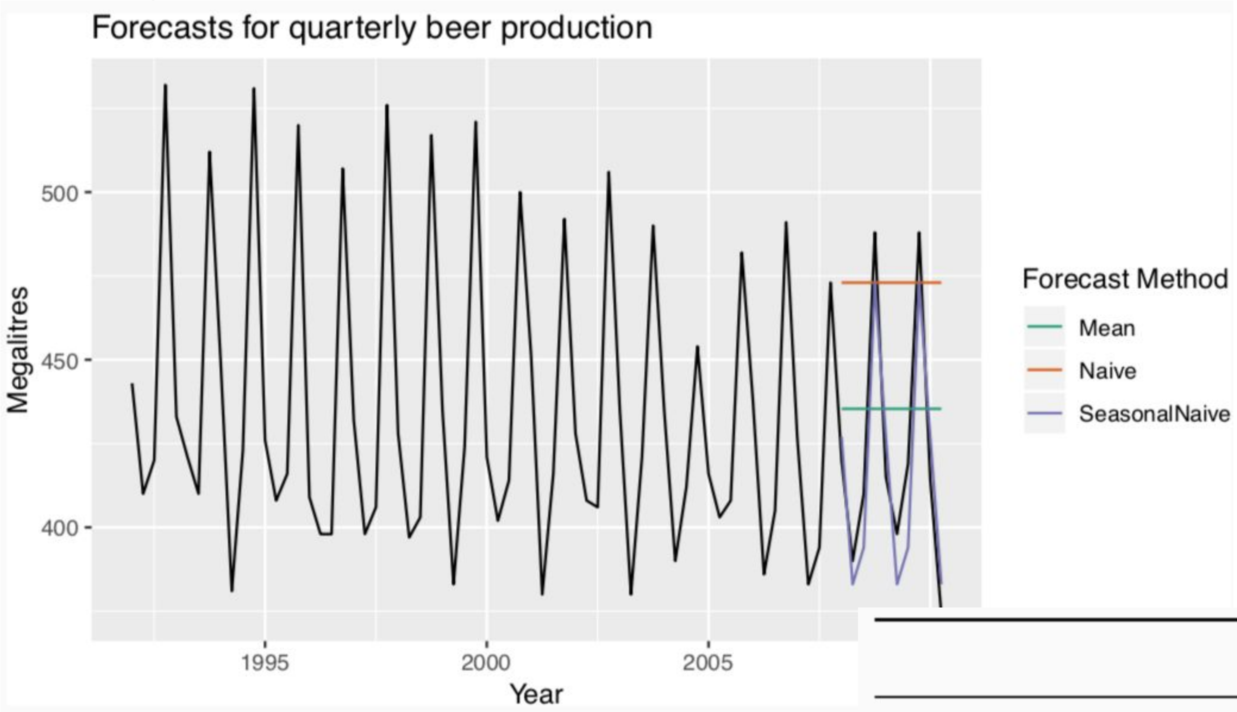
Un **Intervalo de predicción** es un intervalo dentro del cual esperamos que caiga una predicción y_t con determinada probabilidad.

Por ejemplo, si asumimos que los errores están distribuidos normalmente, un intervalo de predicción del 95% es:

$$\hat{y}_{t+h} \pm 1.96\hat{\sigma}_h$$

donde $\hat{\sigma}_h$ es una estimación de la desviación estándar de la distribución de la predicción para el paso h

Cómo medimos la precisión?



	RMSE	MAE	MAPE	MASE
Mean method	38.45	34.83	8.28	2.44
Naïve method	62.69	57.40	14.18	4.01
Seasonal naïve method	14.31	13.40	3.17	0.94

Intervalos de Predicción para predicciones de un paso (h=1)

En general, cuando *h=1*, la desviación estándar de la distribución de las predicciones es *similar* a la desviación estándar de los residuos:

- Si las predicciones no requieren estimación de parámetros, como ser la predicción Naïve, entonces ambas desviaciones estándar coinciden

- En el caso contrario, $\hat{\sigma}_1$ suele ser levemente superior a σ_1, aunque habitualmente esa diferencia es ignorada.

Intervalos de Predicción para Naïve para predicciones de múltiples pasos (h>1)

Cómo obtenemos los intervalos?

Consideramos el método naïve, para multiples pasos (múltiples h). Asumamos que el modelo es

$$Y_t = Y_{t-1} + \epsilon_t$$

donde ε_t es iid con varianza σ^2.

Podemos obtener Y_{t+h} reemplazando los valores correspondientes:

$$Y_{t+h} = Y_{t+h-1} + \epsilon_{t+h} = Y_t + \epsilon_{t+1} + \epsilon_{t+2} + \ldots + \epsilon_{t+h}$$

Intervalos de Predicción para Naïve

Entonces

$$Y_{t+h} = Y_t + \textstyle\sum_{i=1}^{h} \epsilon_{t+i}$$

De esta forma, el punto que predecimos en h es:

$$\begin{aligned} y_{t+h} &= E(Y_{t+h} | y_{1:t}) \\ &= E(Y_t + \sum_{i=1}^{h} \epsilon_{t+i} | y_{1:t}) \\ &= y_t \end{aligned}$$

Intervalos de Predicción para Naïve

La varianza es:

$$
\begin{aligned}
Var(Y_{t+h}|y_{1:t}) &= Var(y_t + \sum_{i=1}^{h} \epsilon_{t+i}|y_{1:t}) \\
&= h\sigma^2
\end{aligned}
$$

Asumiendo que ε_t es normalmente distribuida $N(0, \sigma^2)$:

$$
Y_{t+h}|y_{1:t} \sim N(y_t, h\sigma^2)
$$

Intervalos de Predicción para Naïve

El hecho de ser Y_{t+h} normalmente distribuida, determina que el intervalo de predicción sea:

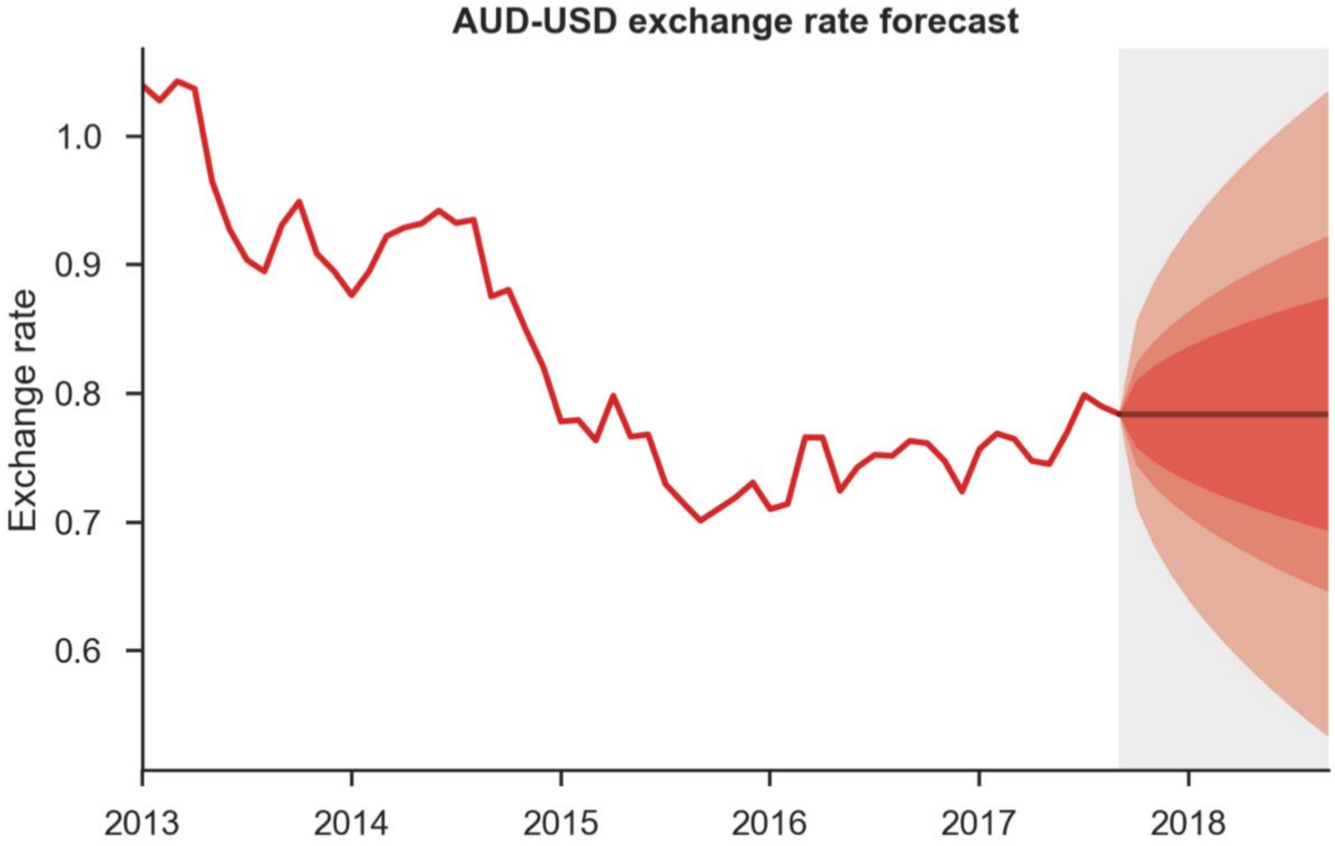
$$y_t \pm z_{\alpha/2}\sqrt{h\hat{\sigma}^2}$$

donde

$$\hat{\sigma}^2 = \frac{\sum_{i=2}^{t} {y_i - y_{i-1}}^2}{t-1}$$

y $z_{\alpha/2}$ es el valor crítico de la distribución normal.

Intervalos de Predicción para Naïve



Regresión

•••

Exponential smoothing

En el modelo de **exponential smoothing**, las predicciones son promedios pesados de observaciones pasadas, donde los pesos decaen exponencialmente cuando nos movemos al pasado:

$$\hat{y}_{t+1} = l_t$$

$$l_t = \alpha y_t + (1-\alpha) l_{t-1}$$

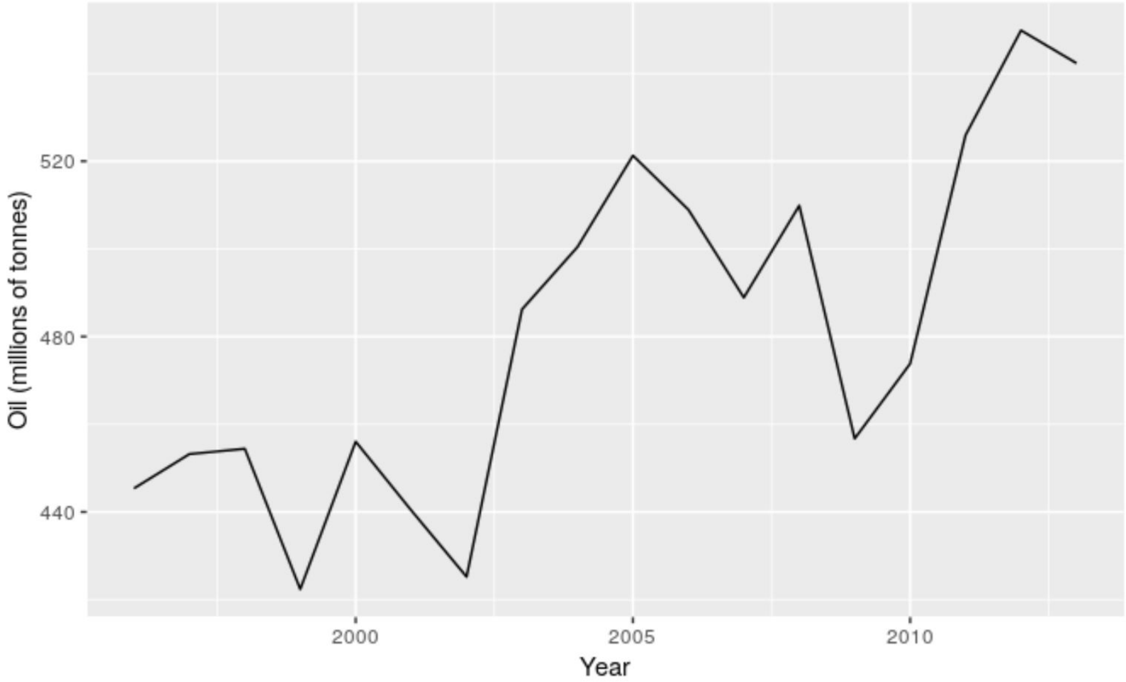
donde asumimos que l_0 está dado y $0 < \alpha < 1$.

Exponential smoothing

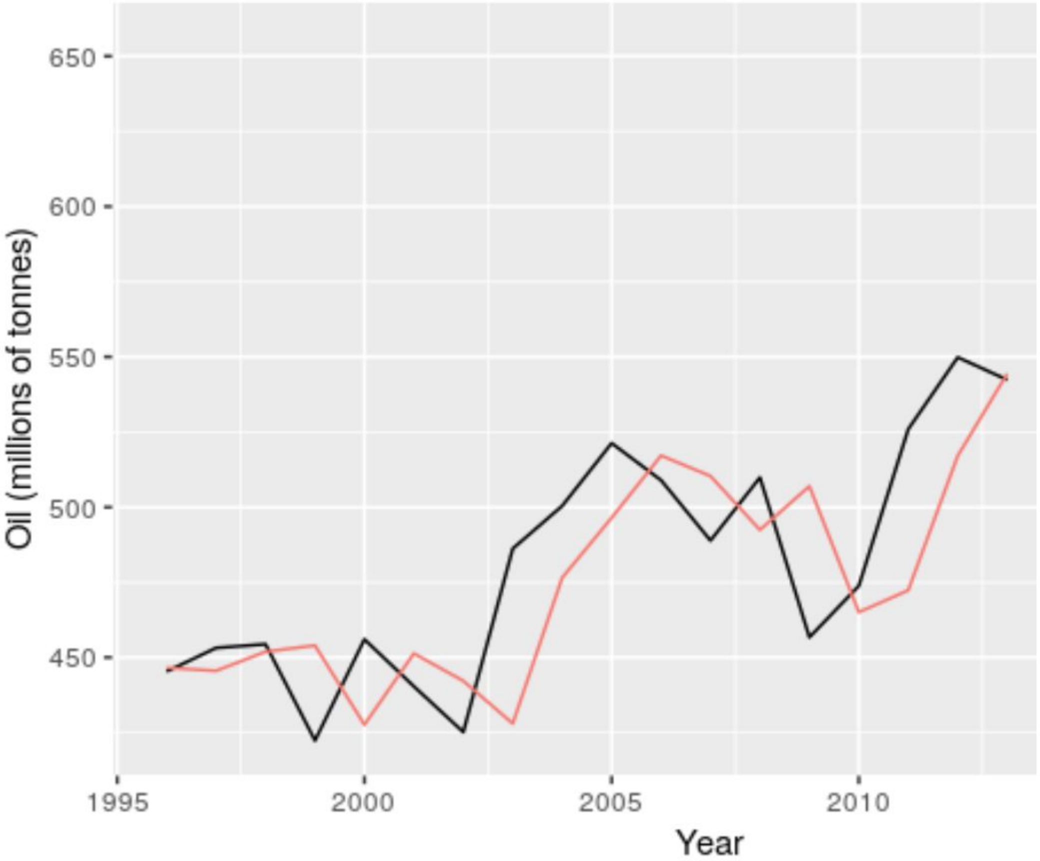
$$
\begin{aligned}
l_1 &= \alpha y_1 + (1-\alpha) l_0 \\
l_2 &= \alpha y_2 + (1-\alpha) l_1 \\
&= \alpha y_2 + (1-\alpha)\alpha y_1 + (1-\alpha)^2 l_0 \\
l_3 &= \alpha y_3 + (1-\alpha) l_2 \\
&= \alpha y_3 + (1-\alpha)\alpha y_2 + (1-\alpha)^2 \alpha y_1 + (1-\alpha)^3 l_0 \\
l_4 &= \alpha y_4 + (1-\alpha) l_3 \\
&= \alpha y_4 + (1-\alpha)\alpha y_3 + (1-\alpha)^2 \alpha y_2 + (1-\alpha)^3 \alpha y_1 + (1-\alpha)^4 l_0
\end{aligned}
$$

Exponential smoothing

El método se suele usar cuando los datos no tienen patrones de tendencia ni estacional claros. Ej



Exponential smoothing



Holt Winters

Holt exponential smoothing

El método de **Holt** (o con corrección de tendencia) incorpora el manejo de la tendencia:

$$
\begin{aligned}
\hat{y}_{t+1} &= l_t + b_t \\
l_t &= \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \\
b_t &= \beta(l_t - l_{t-1}) + (1-\beta) b_{t-1}
\end{aligned}
$$

donde asumimos que l_0 y b_0 están dados y $0 < \alpha < 1$, $0 < \beta < 1$.

Holt winters exponential smoothing

Así como **Holt** incorpora el manejo de la tendencia, **Holt Winters** agrega correcciones de patrones estacionales en los datos:

$$
\begin{aligned}
\hat{y}_{t+1} &= l_t + b_t + S_{t+1-L} \\
l_t &= \alpha(y_t - S_{t-L}) + (1-\alpha)(l_{t-1} + b_{t-1}) \\
b_t &= \beta(l_t - l_{t-1}) + (1-\beta)b_{t-1} \\
S_t &= \delta(y_t - l_t) + (1-\delta)S_{t-L}
\end{aligned}
$$

donde L es la frecuencia estacional, asumimos que l_0, b_0 y S_{i-L} están dados y $0 < \alpha < 1$, $0 < \beta < 1$, $0 < \delta < 1$.

ARIMA

•••

ARIMA

Los modelos **ARIMA** intentan describir la autocorrelación en los datos. Estos modelos suelen ser complementarios a exponential smoothing.

ARIMA es (o era) uno de los modelos más usados en la predicción de series temporales.

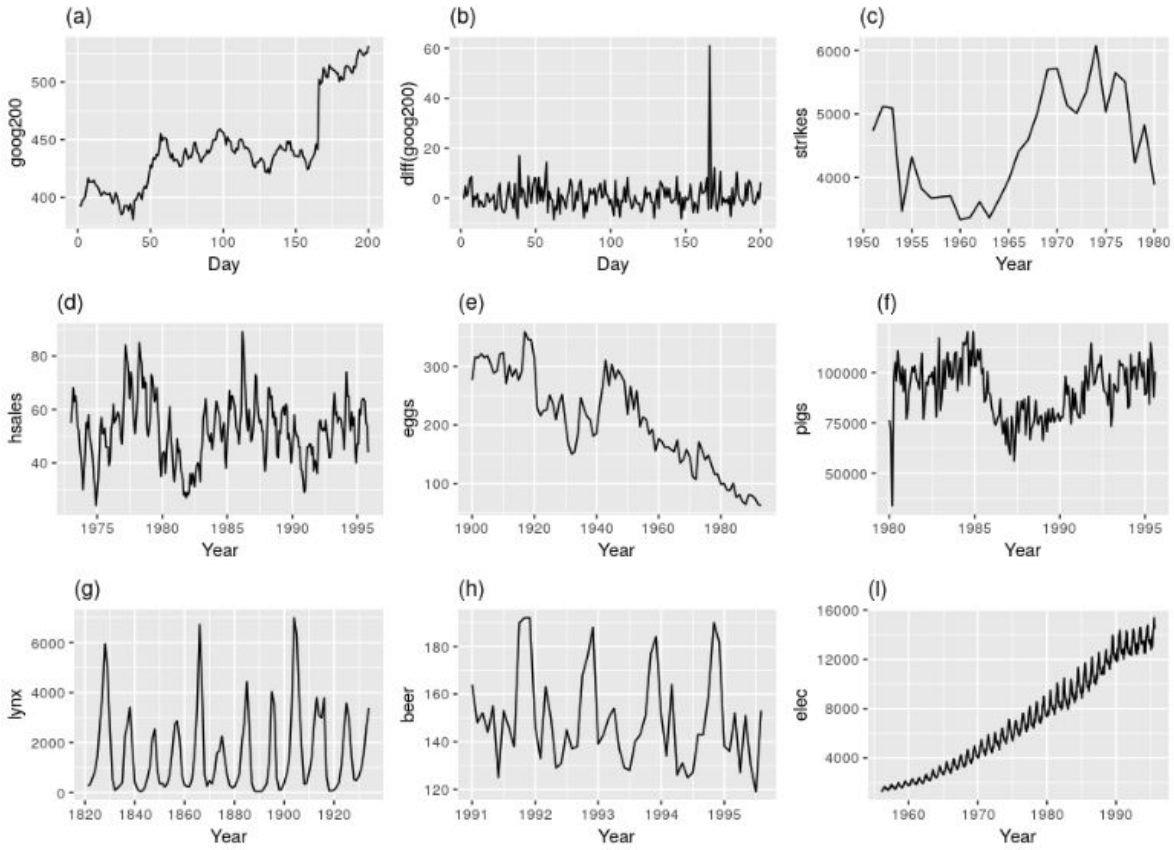
Para introducir ARIMA, debemos revisar nuevos conceptos: stationarity y diferenciación.

Estacionariedad (Stationarity)

Una serie temporal con **estacionariedad** es aquella cuyas propiedades no dependen del tiempo en el cual los datos son observados.

Por ejemplo, las series de "white noise" tienen estacionariedad: no importa el instante en que se observen, lucen bastante parecidas en cualquier momento. Por su parte, las series con tendencia o estacionales no tienen estacionariedad.

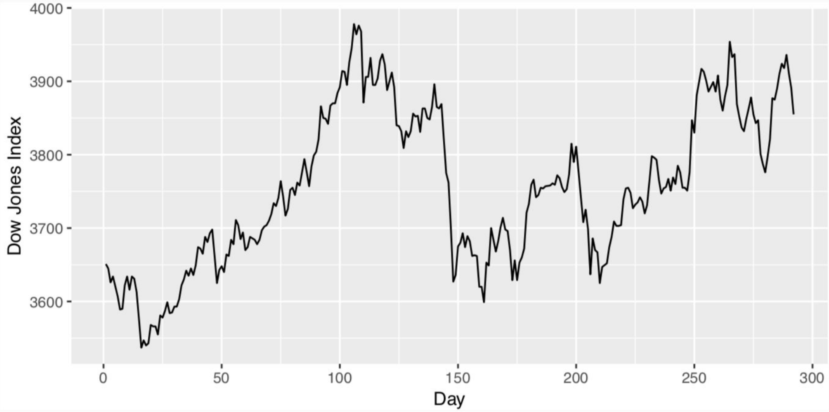
Estacionariedad (Stationarity)

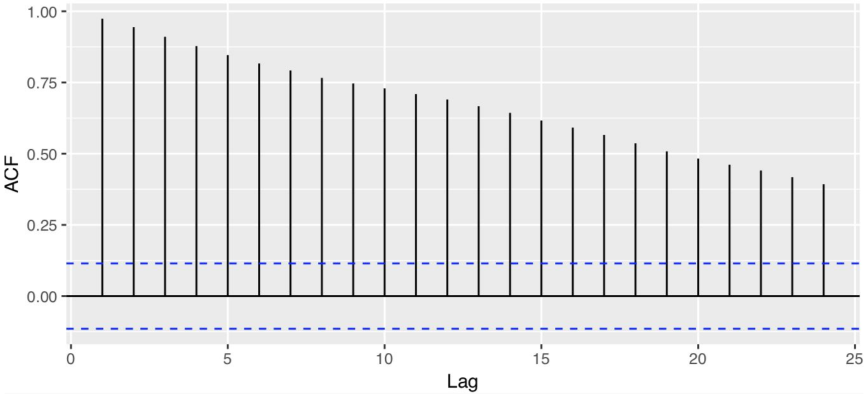


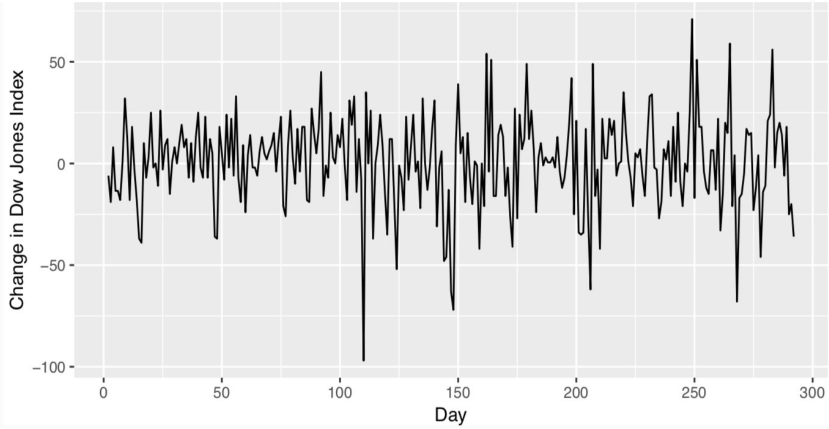
Cómo identificamos estacionariedad?

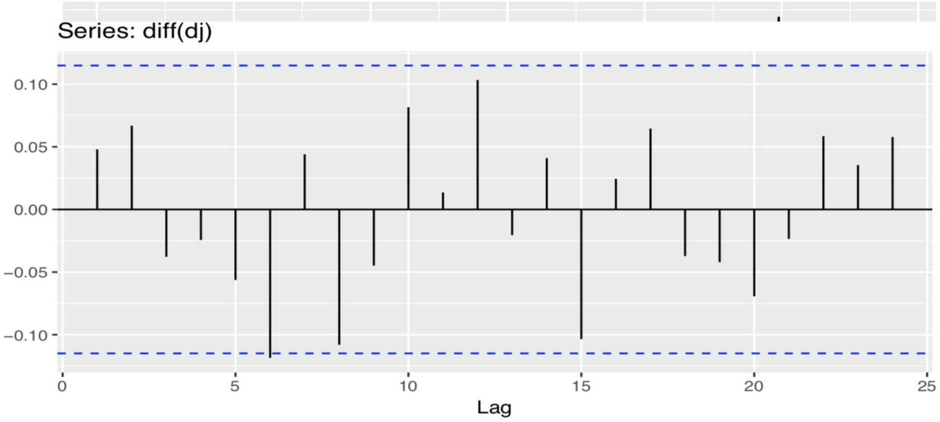
- time plot
- ACF: el ACF de las series con estacionariedad caen a cero relativamente rápido
- r_1 suele ser grande y positivo en las series sin estacionariedad

Cómo identificamos estacionariedad?









Diferenciales

Diferencial de primer orden

$$\Delta Y_t = Y_t' = Y_t - Y_{t-1}$$

Ejemplo:

Para el modelo Naïve, tenemos que

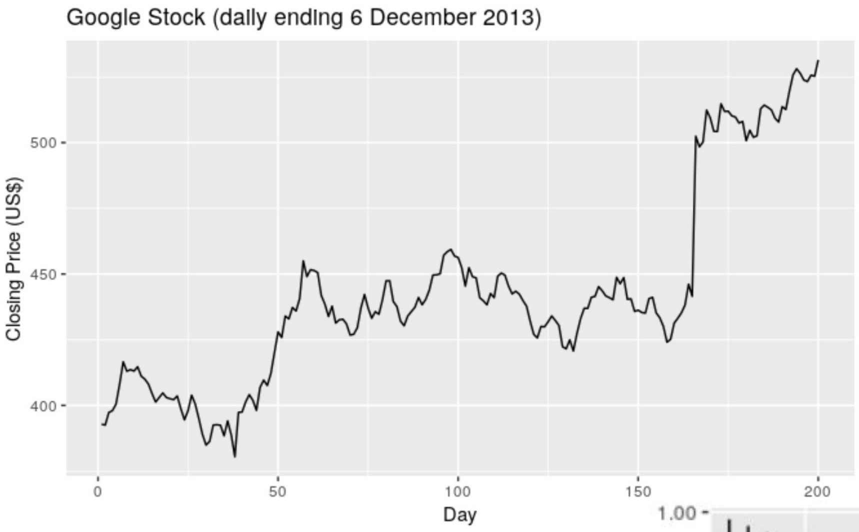
$$Y_t = Y_{t-1} + \epsilon_t$$

por lo tanto, la primera diferencia representa sólo white noise

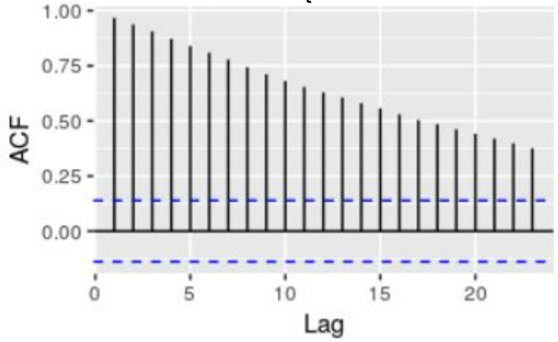
$$\Delta Y_t = Y_t - Y_{t-1} = \epsilon_t$$

(la diferencia estabiliza la media de una serie temporal quitando los los fenómenos de tendencia y estacionalidad).

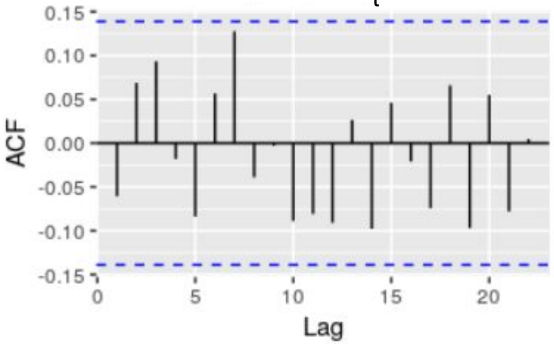
Diferenciales



Y_t



ΔY_t



Diferenciales

Diferencial de segundo orden

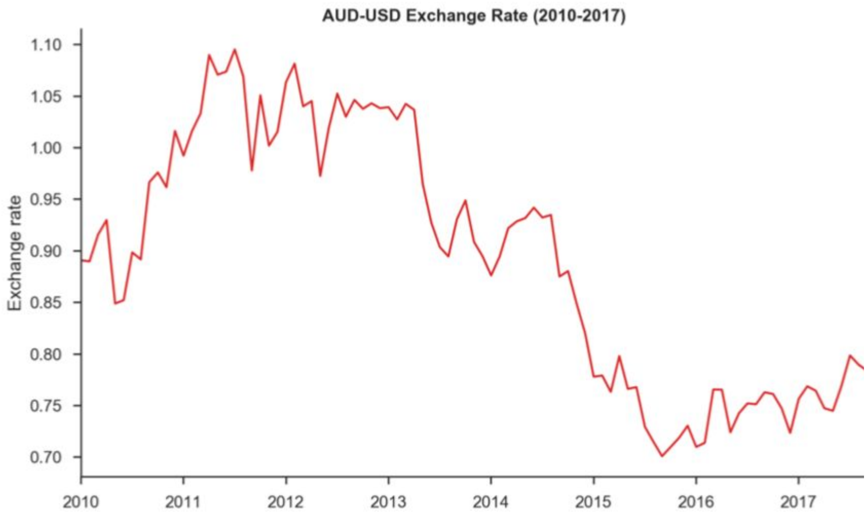
En algunos casos extraordinarios es necesario diferenciar la serie dos veces para obtener estacionariedad:

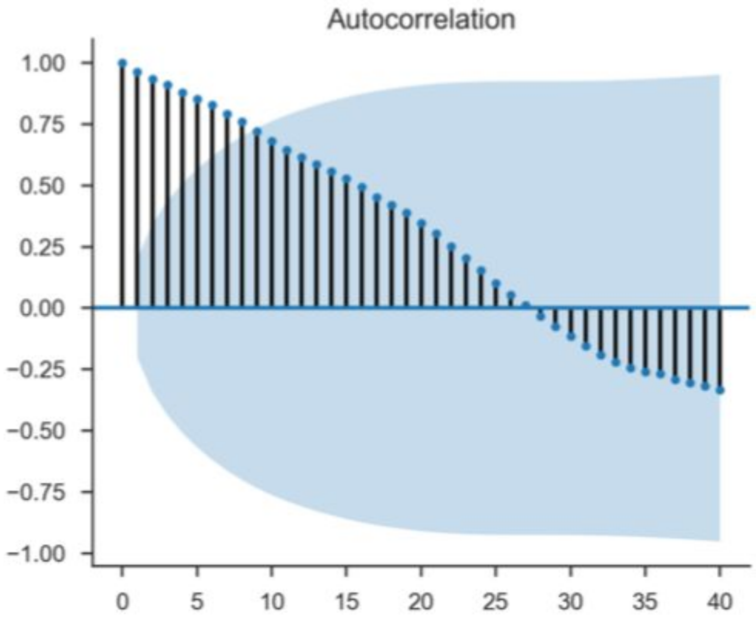
$$\Delta^2 Y_t = Y_t'' = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) = Y_t - 2Y_{t-1} + Y_{t-2}$$

El ACF (como en el ejemplo anterior) nos ayuda a determinar si la serie temporal necesita ser diferenciada:

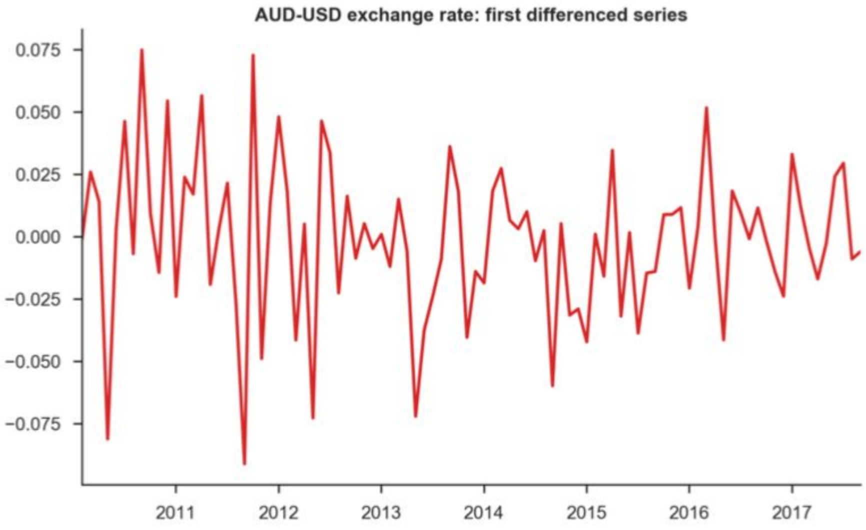
- ACF de una serie no estacionaria decrece lentamente
- ACF de una serie estacionaria cae a cero rápidamente

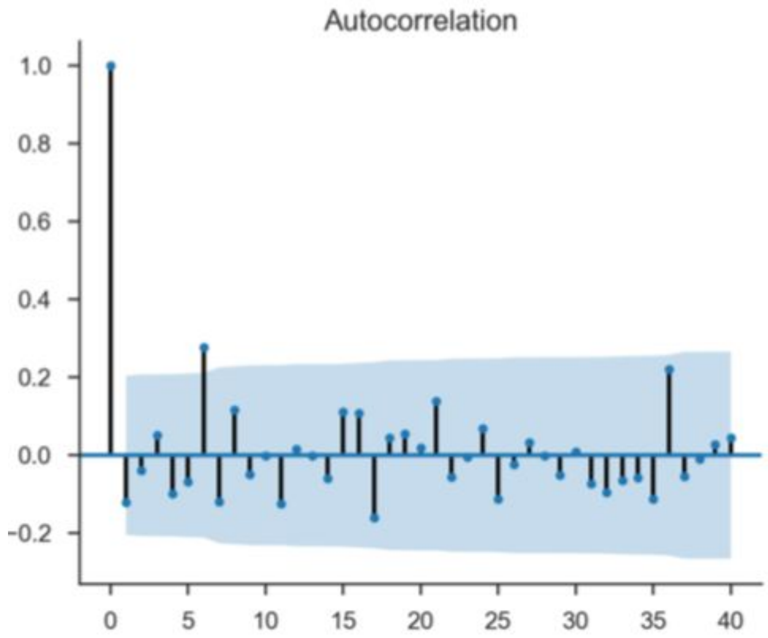
Diferenciales





Diferenciales





Diferenciales

Diferencial estacional
Para los casos en que el fenómeno sea estacional, se usa el diferencial estacional:

$$\Delta^m Y_t = Y_t - Y_{t-m}$$

donde *m* es la frecuencia de las estaciones.

- Las series con patrones de tendencia y estacional, suelen requerir diferenciación primaria y estacional:

$$\Delta^m(\Delta Y_t) = (Y_t - Y_{t-1}) - (Y_{t-m} - Y_{t-m-1})$$

Diferenciales

Diferencial estacional y primera diferencial

- no hay diferencia en el orden en que se aplican
- si la serie tiene clara estacionalidad, se suele aplicar diferencial estacional primero porque muchas veces la serie resultante suele ser estacionaria.

En general, si se diferencia, las diferenciales deben ser **interpretables!**

- la primera diferencial son cambios entre una observación y la siguiente
- diferenciales estacionales suelen ser el cambio entre un año y el siguiente

Notación Backshift

(aka Backward shift)

$$BY_t = Y_{t-1}$$

Se suelen aplicar multiples operadores Backward:

$$B(BY_t) = B^2Y_t = Y_{t-2}$$

O incluso:

$$B^{12}Y_t = Y_{t-12}$$

Notación Backshift

Esta notación se suele usar para describir la **diferenciación**:

$$Y'_t = Y_t - Y_{t-1} = Y_t - BY_t = (1 - B)Y_t$$

$$Y''_t = Y_t - 2Y_{t-1} + Y_{t-2} = (1 - B)^2 Y_t$$

Diferencial estacional seguida por la primera diferencia se denota:

$$(1 - B)(1 - B^m)Y_t$$

Modelo Autorregresivo (AR)

El **Modelo Autorregresivo de orden p AR(**p**)** se define como:

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t$$

donde c es constante, Φ es el parámetro autorregresivo y ε_t es ruido.

Casos especiales de AR(1):

- $\Phi_1 = 0$
- $\Phi_1 = 1$ y $c = 0$
- $\Phi_1 = 1$ y c != 0

Medias Móviles (Moving Averages MA)

El **modelo de Medias Móviles** de orden q **MA(**q**)** se define como:

$$Y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

donde c es constante, θ es el parámetro autorregresivo y ε_t es ruido.

Medias Móviles (Moving Averages MA)

El **modelo de Medias Móviles** de orden q **MA(**q**)** se define como:

$$Y_t = c + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots + \theta_q \epsilon_{t-q}$$

donde c es constante, θ es el parámetro autorregresivo y ε_t es ruido.

Es posible escribir AR(p) como un proceso MA(∞):

$$\begin{aligned} Y_t &= \phi_1 Y_{t-1} + \epsilon_t \\ &= \phi_1 (\phi_2 Y_{t-2} + \epsilon_{t-1}) + \epsilon_t \\ &= \phi_1^2 Y_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t \\ &= \phi_1^3 Y_{t-3} + \phi_1^2 \epsilon_{t-2} + \phi_1 \epsilon_{t-1} + \epsilon_t \\ &= \ldots \end{aligned}$$

Invertibilidad

En el caso en que MA(q) se pueda escribir como AR(∞), decimos que el modelo MA es **invertible**. Tal invertibilidad siempre se puede lograr si se agregan condiciones en los parámetros del modelo MA:

Los ceros complejos de

$$1 + \phi_1 z + \phi_2 z^2 + \ldots + \phi_q z^q$$

caen fuera del círculo unitario.

ARIMA

El modelo AutoRegresivo con Medias Móviles (**ARMA**) se define por:

$$\begin{aligned} Y_t = c &+ \phi_1 Y_{t-1} + \ldots + \phi_p Y_{t-p} \\ &+ \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} \\ &+ \epsilon_t \end{aligned}$$

ARMA incluye predicciones con valores lagged y errores lagged.

El modelo AutoRegresivo Integrado con Medias Móviles (**ARIMA**) está dado por:

$(1 - B)^d Y_t$ seguido por un modelo ARMA

ARIMA (Notación)

ARIMA(p, d, q):

- AR(p)
- I(d) donde d es el grado de la diferenciabilidad estacional
- MA(q)

A qué modelo corresponden los siguientes?

- ARIMA(0,0,0)
- ARIMA(0,1,0)
- ARIMA(p,0,0)
- ARIMA(0,0,q)

ARMA en Backshift

$$
\begin{aligned}
Y_t = c &+ \phi_1 B Y_t + \ldots + \phi_p B^p Y_t \\
&+ \theta_1 B \epsilon_t + \ldots + \theta_q B^q \epsilon_t \\
&+ \epsilon_t
\end{aligned}
$$

O:

$$(1 - \textstyle\sum_{i=1}^{p} \phi_i B^i) Y_t = c + (1 + \textstyle\sum_{i=1}^{q} \theta_i B^i) \epsilon_t$$

Por ejemplo, para ARIMA(1,1,1):

$$(1 - \phi_1 B)(1 - B) Y_t = c + (1 + \theta_1 B) \epsilon_t$$

ARIMA estacional

Los modelos ARIMA estacionales se construyen de la siguiente manera:

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

Donde (p,d,q) corresponden al ARIMA (no estacional) y D es el orden de diferencial estacional, y P y Q son los órdenes de las componentes estacionales AR y MA; m es la frecuencia de estaciones.

ARIMA estacional

Ejemplo: **ARIMA(1,1,1)(1,1,0)$_{12}$**:

ARIMA(1,1,1): $(1-\phi_1 B)(1-B)Y_t = c + (1+\theta_1 B)\epsilon_t$

ARIMA(1,1,1)(1,1,0)$_{12}$:

$$(1-\phi_1 B)(1-\phi_2 B^{12})(1-B)(1-B^{12})Y_t = c + (1+\theta_1 B)\epsilon_t$$

Series Temporales Multivariadas

...

Series Multivariadas

Time	Temperature
5:00 am	59 °F
6:00 am	59 °F
7:00 am	58 °F
8:00 am	58 °F
9:00 am	60 °F
10:00 am	62 °F
11:00 am	64 °F
12:00 pm	66 °F
1:00 pm	67 °F
2:00 pm	69 °F
3:00 pm	71 °F
4:00 pm	71 °F
5:00 pm	71 °F
6:00 pm	69 °F
7:00 pm	68 °F
8:00 pm	65 °F
9:00 pm	64 °F

Ref https://www.analyticsvidhya.com/blog/2018/09/multivariate-time-series-guide-forecasting-modeling-python-codes/

Series Multivariadas

Time	Temperature	cloud cover	dew point	humidity	wind
5:00 am	59 °F	97%	51 °F	74%	8 mph SSE
6:00 am	59 °F	89%	51 °F	75%	8 mph SSE
7:00 am	58 °F	79%	51 °F	76%	7 mph SSE
8:00 am	58 °F	74%	51 °F	77%	7 mph S
9:00 am	60 °F	74%	51 °F	74%	7 mph S
10:00 am	62 °F	74%	52 °F	70%	8 mph S
11:00 am	64 °F	76%	52 °F	65%	8 mph SSW
12:00 pm	66 °F	80%	52 °F	60%	8 mph SSW
1:00 pm	67 °F	78%	52 °F	58%	10 mph SW
2:00 pm	69 °F	71%	52 °F	54%	10 mph SW
3:00 pm	71 °F	75%	52 °F	52%	11 mph SW
4:00 pm	71 °F	78%	52 °F	52%	11 mph SW
5:00 pm	71 °F	78%	52 °F	52%	12 mph SW
6:00 pm	69 °F	78%	52 °F	54%	11 mph SW
7:00 pm	68 °F	87%	53 °F	60%	12 mph SW
8:00 pm	65 °F	100%	54 °F	66%	11 mph SSW
9:00 pm	64 °F	100%	55 °F	72%	13 mph SSW

Ref https://www.analyticsvidhya.com/blog/2018/09/multivariate-time-series-guide-forecasting-modeling-python-codes/

Series Multivariadas

Modelos Autorregresivos (AR):

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t$$

Cómo lo adaptamos a dos variables?

Series Multivariadas

Modelos Autorregresivos (AR):

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t$$

Cómo lo adaptamos a dos variables?

$$\begin{aligned} Y_{1,t} = c &+ \phi_{11} Y_{1,t-1} + \phi_{21} Y_{1,t-2} + \ldots + \phi_{p1} Y_{1,t-p} \\ &+ \phi_{21} Y_{2,t-1} + \phi_{22} Y_{2,t-2} + \ldots + \phi_{2p} Y_{2,t-p} \\ &+ \epsilon_t \end{aligned}$$

O, gráficamente: pizarrón

Series Multivariadas

Modelo Autorregresivo Vectorial (VAR):

$$Y_t = c + A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots + A_p Y_{t-p} + \epsilon_t$$

ARMA Vectorial (VARMA):

$$\begin{aligned} Y_t = c &+ A_1 Y_{t-1} + \ldots + A_p Y_{t-p} \\ &+ \theta_1 \epsilon_{t-1} + \ldots + \theta_q \epsilon_{t-q} \\ &+ \epsilon_t \end{aligned}$$

Detección de Anomalías

Anomalías

El problema de detección de anomalías para series temporales se suele definir como la detección de "outliers" relativos a una señal.

Tipos de anomalías:

- **Outliers aditivos:** variaciones inesperadas prácticamente instantáneas

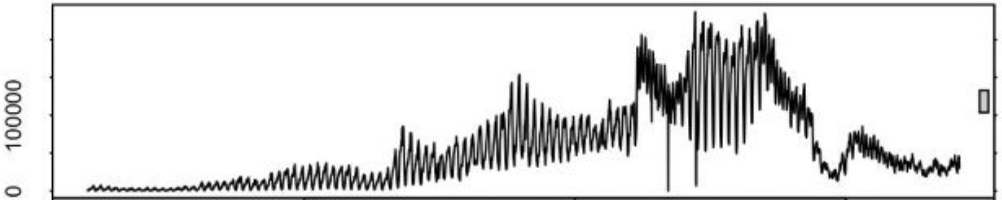
- **Cambios temporales:** variaciones inesperadas en un período corto de tiempo

- **Cambios de nivel (level shifts):** variaciones que no modifican la forma de la señal, sino sus niveles, durante un período de tiempo

Ref **https://blog.statsbot.co/time-series-anomaly-detection-algorithms-1cef5519aef2**

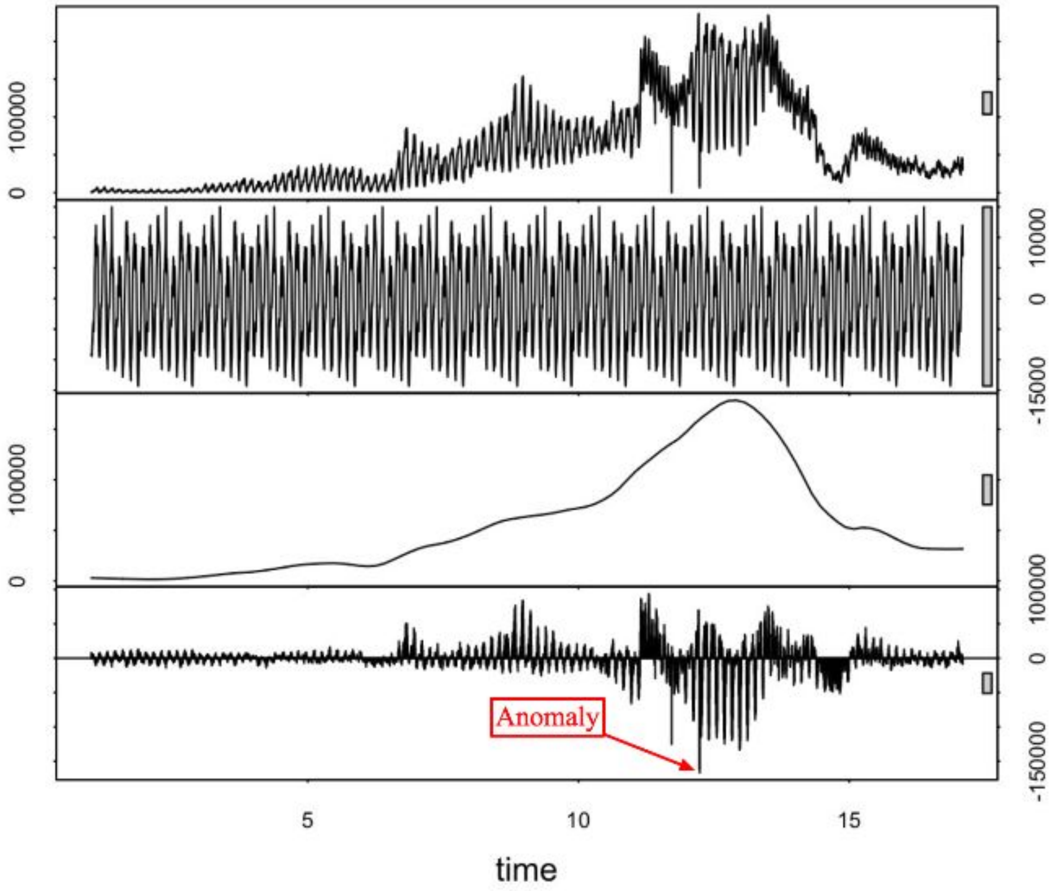
Detección de Anomalías

STL



Detección de Anomalías

STL



Detección de Anomalías

STL:

Librería más famosa: https://github.com/twitter/AnomalyDetection

Ventajas:

- Simple
- Robusto
- Intuitivo

Desventajas:

- Bastante rígido a la hora de tunearlo. Lo único que se puede variar son los intervalos de confidencia

Detección de Anomalías

CART (Classification and Regression Trees)

- **Supervisado:** se entrenan árboles con datos etiquetados como anómalos/no-anómalos.

- **No Supervisado:** se usa CART para predecir los próximos puntos en la serie, y un intervalo de confianza determina las anomalías

Ventajas:
- ?No está sujeto a la "forma" de la señal

Desventajas:
- Cómo seleccionamos las features?

Detección de Anomalías

- **ARIMA**
- **Exponential Smoothing**
- **Redes Neuronales**

LSTM

•••

LSTMs

- Introducción a LSTM: http://web.cs.ucla.edu/~kwchang/teaching/NLP16/slides/NN.pdf
- Seguimos el ejemplo de: https://www.altumintelligence.com/articles/a/Time-Series-Prediction-Using-LSTM-Deep-Neural-Networks
- Código: https://github.com/jaungiers/LSTM-Neural-Network-for-Time-Series-Prediction
- Docker: https://hub.docker.com/r/jupyter/tensorflow-notebook/

Otras herramientas

•••

Otras herramientas

- **R's forecast**
- **Facebook Prophet**
  - https://facebook.github.io/prophet/
- **Sktime -** A scikit-learn for time series data - **new 2019** v0.3
  - https://github.com/alan-turing-institute/sktime
- **Amazon Forecast**
  - https://aws.amazon.com/blogs/aws/amazon-forecast-time-series-forecasting-made-easy/
- **TSDBs (Time series databases)**
  - **InfluxDB, Prometheus, OpenTSDB, Timescale**
- **Awesome lists in github**
  - awesome_time_series_in_python
  - awesome-TS-anomaly-detection
  - awesome-time-series-database

Apéndice

Proceso de desestacionalización

Vimos que las series se pueden descomponer en

$$Y_t = T_t + S_t + C_t + E_t$$

donde cada componente representa a cada patrón (aditiva).

Los métodos **X11** (60s) y **ARIMA** (70s) permitieron que las herramientas de desestacionalización progresaran en dos direcciones:

- Enfoque no paramétrico (empírico)
- Enfoque paramétrico (basado en modelos)

Proceso de desestacionalización

Enfoque no paramétrico (empírico)

Estima las componentes no observadas sin implementar un modelos estadístico; usualmente las componentes se estiman mediante la aplicación de filtros lineales sucesivamente, los cuales pueden ser considerados como regresiones locales en ventanas móviles. La metodología más usada es **X12-ARIMA**.

Enfoque paramétrico (basado en modelos)

Explicita un modelo de la serie o de sus componentes. Puede que tales modelos sean determinísticos o bien estocásticos. La metodología más usada es **TRAMO-SEATS**

Proceso de desestacionalización

X12-ARIMA

- Se eliminan automáticamente los valores identificados como "outliers". Se identifican y "tratan" los efectos de días laborales, longitud del mes, efecto de feriados como Pascuas, Carnaval.

- Se extiende la serie anterior con predicciones generadas a partir de 5 modelos ARIMA, aplicados secuencialmente.

- Se aplican a la serie anterior el método de medias móviles (MA: medias móviles) y Henderson (X11) para obtener las componentes estacional, tendencia/cíclica e irregular.

Proceso de desestacionalización

TRAMO-SEATS

Se encarga de la corrección de la variación estacional a corto plazo.

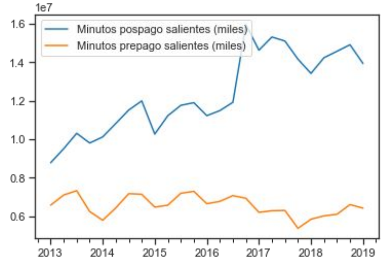
Lo hace en 2 etapas:

- Se detectan y eliminan los "outliers" y "perturbaciones" (que suelen afectar mucho a TRAMO)

- SEATS descompone la serie anterior según un modelo determinado por TRAMO en componentes mediante la estimación y descomposición de la densidad espectral de ARIMA.

Laboratorios

Lab 1 - Series temporales

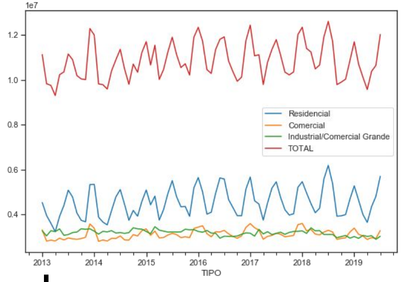


1) Levantar los datos desde el archivo dataset/Telefonia_movil.csv. Los mismos fueron extraídos de http://datosabiertos.enacom.gob.ar/visualizations/29890/trafico-de-suscriptores-de-telefonia-movil-miles-de-minutos/

2) Explorar los datos, visualizarlos

3) Implementar dos modelos (como ser Random Walk (Naïve) y Exponential Smoothing)

4) Visualizar los resultados y evaluarlos, comparando ambas implementaciones



Lab 2 - Series con estacionalidad

1) Levantar los datos desde el archivo dataset/*demanda.csv*. Los mismos fueron extraídos de los datos del informe mensual en http://portalweb.cammesa.com/memnet1/Pages/descargas.aspx

2) Explorar los datos y visualizar descomposición de datos

3) Modelar con Holt Winters Smoothing con diferentes parámetros

4) Visualizar los resultados y evaluar su poder predictivo, comparando ambos modelos