



Universidad **Ricardo Palma**

RECTORADO

PROGRAMA DE ESPECIALIZACIÓN EN CIENCIA DE DATOS

Formamos seres humanos para una cultura de paz

PROGRAMA DE ESPECIALIZACIÓN DATA SCIENCE NIVEL I

R + Python



MÓDULO I



python™

**Modelos de Series Temporales :
Enfoque Descomposición,
Enfoque Dominio del Tiempo**



A nuestro recordado Maestro

Dr. Erwin Kraenau Espinal, Presidente de la Comisión de Creación de la Maestría en Ciencia de los Datos



PROGRAMA DE ESPECIALIZACIÓN EN DATA SCIENCE NIVEL I

« Para poder **seguir** a veces hay que **empezar de nuevo**»



Agenda

- ✓ Proceso Estacionario.
- ✓ Características de los Procesos Estacionarios.
- ✓ Definiciones en Series de Tiempo.
- ✓ Enfoques en el Análisis de Series de Tiempo.
- ✓ Enfoque de Descomposición. Modelos Basados en Suavización Exponencial.
- ✓ Enfoque de Dominio del Tiempo o Modelos de Box – Jenkins.
- ✓ Modelos ARIMA , SARIMA.
- ✓ Análisis de Series de Tiempo.
- ✓ Aplicaciones de Series de Tiempo.



Series temporales paramétricas i :

Procesos Estocástico

- Un proceso estocástico es una familia de variables aleatorias definidas en un espacio de probabilidad. $(\Omega, \mathcal{A}, \mathbb{P})$.
- Una serie de tiempo es una realización de un proceso, en forma secuencial y en intervalos de tiempo, que tiene un inicio (t_0) y un fin (t_n).
- Características de un Proceso Estocástico:
 - ✓ **Gaussiano**
 - ✓ **Estacionario**
 - ✓ **No Markoviano**



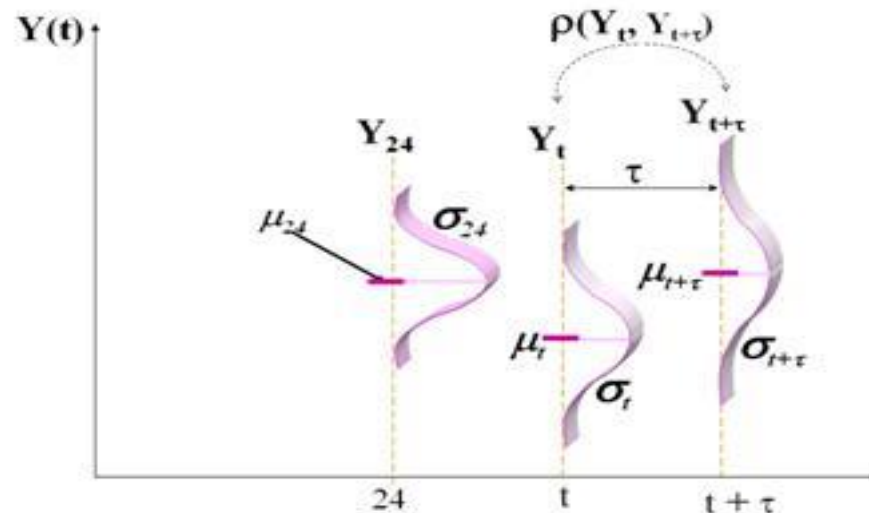
CARACTERÍSTICAS DE UN PROCESO ESTOCÁSTICO

- GAUSSIANO
- ESTACIONARIO
 - ESTACIONARIEDAD EN SENTIDO DÉBIL
 - ESTACIONARIEDAD EN SENTIDO ESTRICTO
- NO MARKOVIANO



Procesos estocásticos

- **Definición:** Un proceso estocástico es una sucesión de variables aleatorias ordenadas en el tiempo (en el caso de series temporales).
- **Definición:** Una serie temporal es una realización del proceso estadístico, es decir, es una observación de T variables aleatorias ordenadas en el tiempo.



Series Temporales Paramétricas

- Una *serie de tiempo* es una colección o conjunto de mediciones de cierto fenómeno o experimento registrados secuencialmente en el tiempo, en forma *equiespaciada* (a intervalos de tiempo iguales).
- Las observaciones de una serie de tiempo serán denotadas por

$$Y(t_1), Y(t_2), \dots, Y(t_n)$$

donde $Y(t_i)$ es el valor tomado por el proceso en el instante t_i .



Series Temporales Paramétricas

- Una de las motivaciones para el estudio del tema surge desde tiempos remotos donde una de las principales inquietudes del hombre ha **sido estimar el futuro** utilizando información del **presente y del pasado**.

Esto se llama *predecir*.

- Es evidente que las diversas instituciones requieren conocer el **comportamiento futuro de ciertos fenómenos con el fin de planificar, preveer o prevenir**.



Series Temporales Paramétricas

- La Estadística ha desarrollado teoría y métodos que apuntan a resolver el **problema de predicción**.
- Sin embargo, este no puede ser resuelto por argumentos puramente matemáticos, debe ser el resultado de una **combinación matemático-especialista**.
- La **predicción es una Ciencia y es un Arte**, y la mayor dificultad es la mala comunicación entre **los analistas de información y de predicción y los usuarios de éstas**.



ENFOQUES PARA EL ANÁLISIS DE SERIES DE TIEMPO

- Los métodos mas utilizados para el análisis de series temporales son:
 - **Enfoque de Descomposición**
 - Enfoque de Dominio del Tiempo o Modelos de Box – Jenkins
 - Enfoque de Dominio de Frecuencias.



Proceso: metodología Análisis series de tiempo

1. Identificación tentativa del modelo



2. Estimación de los parámetros del modelo



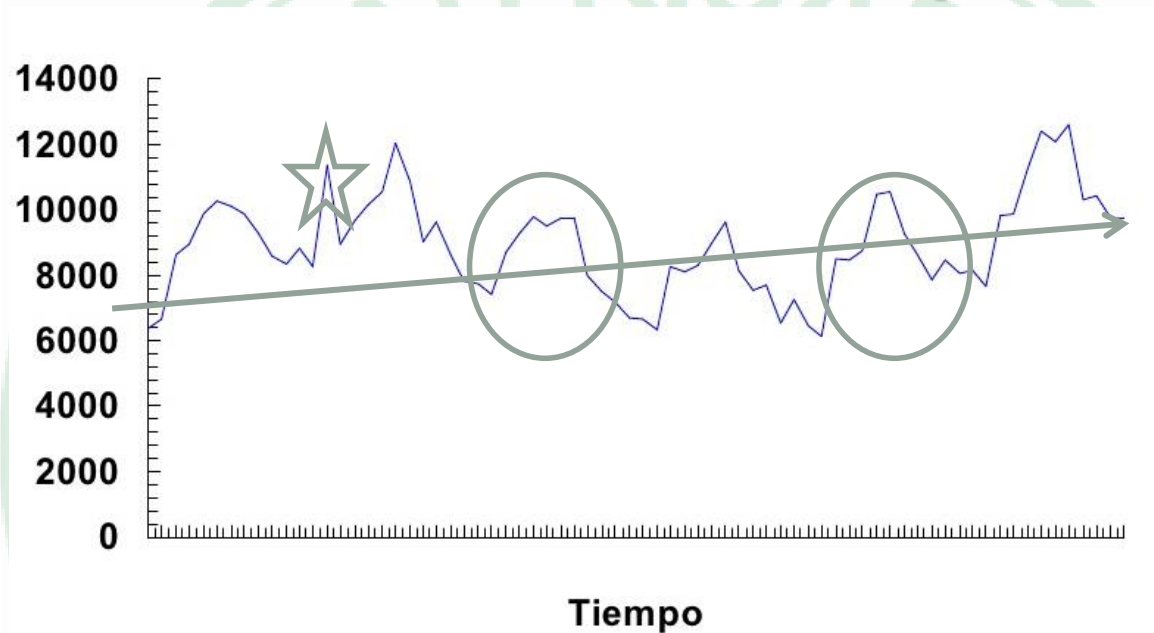
3. Evaluación de diagnósticos para comprobar si el modelo es adecuado; mejorar el modelo si es necesario.



4. Generación de Pronósticos



Series Temporales Paramétricas : Análisis Gráfico de una Serie de Tiempo



Tendencia

Estacionalidad

Aleatoriedad



Series Temporales Paramétricas

- El gráfico de la serie permitirá detectar los siguientes elementos:

a) Outliers:

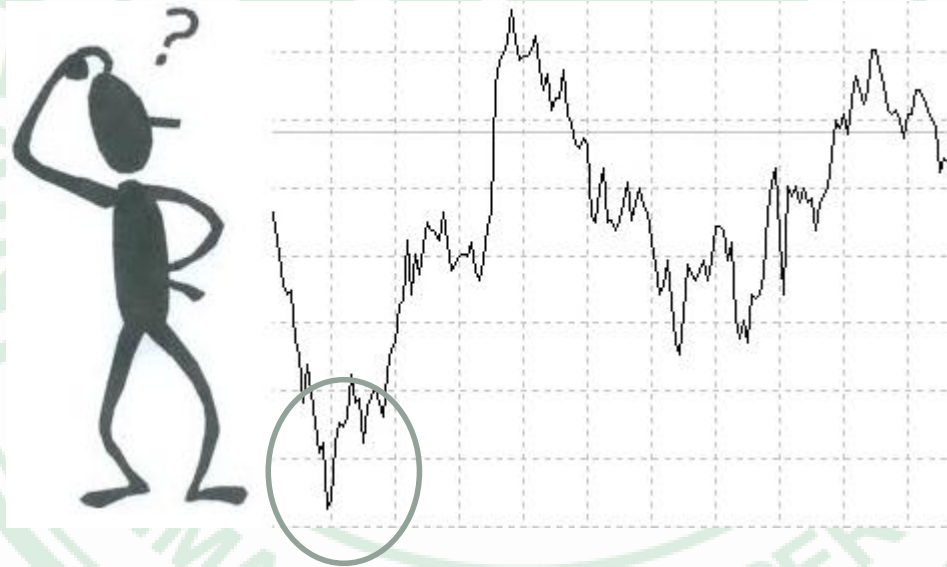
- Se refiere a puntos de la serie que se escapan de lo normal.
- Si se sospecha que una observación es un outliers, se debe reunir información adicional sobre posibles factores que afectaron el proceso.



Series Temporales Paramétricas

➤ Por ejemplo, en un estudio de la producción diaria de cobre se presentó la siguiente situación:

➤ **Outliers**

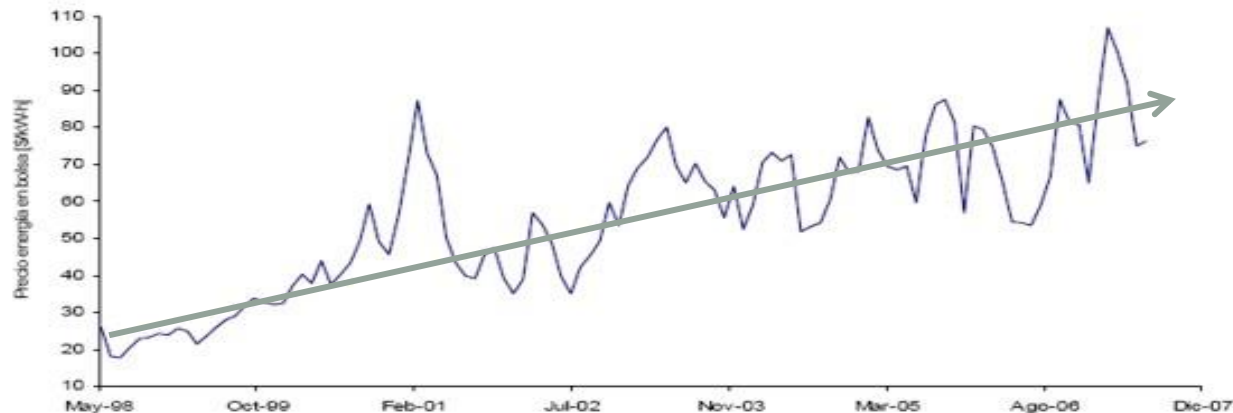


Series Temporales Paramétricas

b) Tendencias:

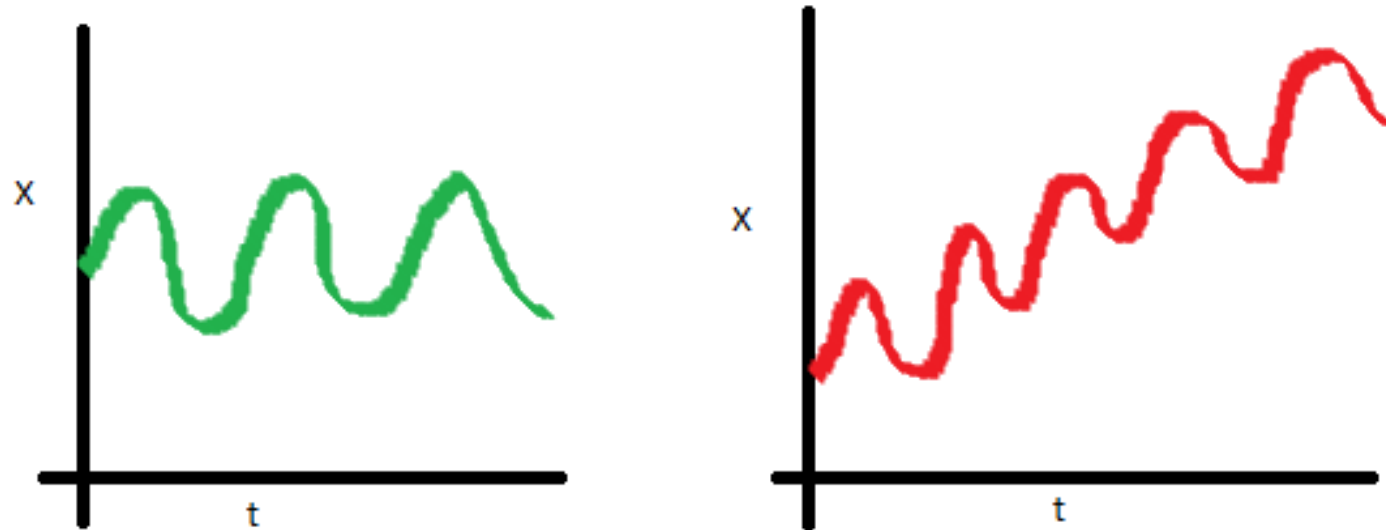
- La tendencia representa el comportamiento predominante de la serie.
- Esta puede ser definida como el cambio de la media a lo largo de un extenso período de tiempo.

GRÁFICA 4
PRECIO DE BOLSA HISTÓRICO MENSUAL [\$/KWH]



Series Temporales Paramétricas

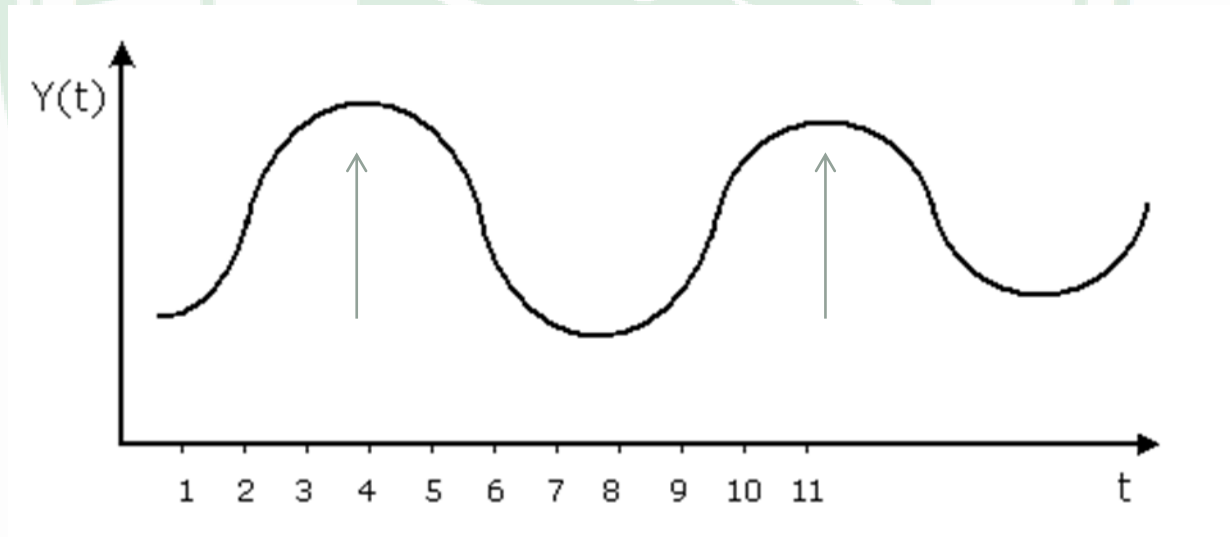
¿ Observa Tendencia en el Gráfico ?



Series Temporales Paramétricas

c) Variaciones cíclicas o estacionales:

- La variación estacional representa un movimiento *periódico* de la serie de tiempo.
- La duración del período puede ser un año, un trimestre, un mes, un día, etc.



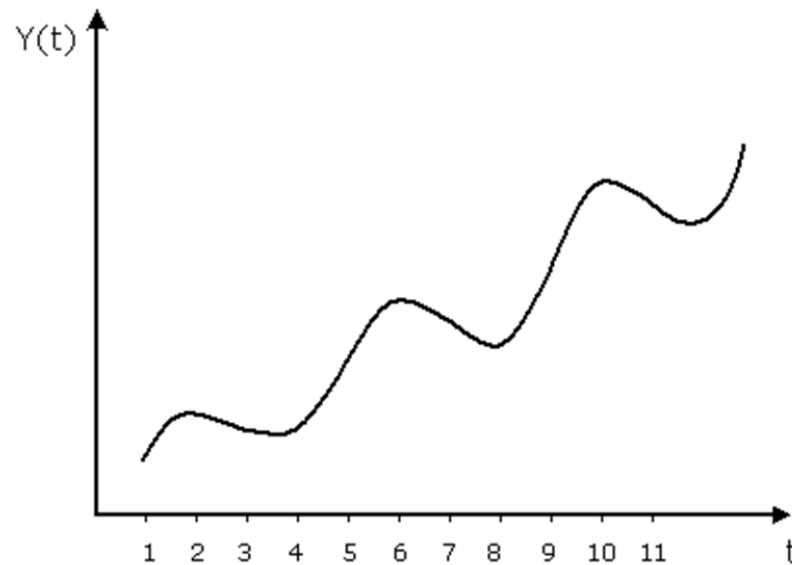
Series Temporales Paramétricas

- Se suele hacer una distinción entre componentes **cíclicas** y **estacionarias**.
- Estas últimas ocurren con períodos identificables, como la estacionalidad del empleo, o de la venta de ciertos productos, cuyo período es un año.
- El término **variación cíclica** se suele referir a ciclos grandes, cuyo período no es atribuible a alguna causa.
- Por ejemplo, fenómenos climáticos, que tienen ciclos que duran varios años.



Series Temporales Paramétricas

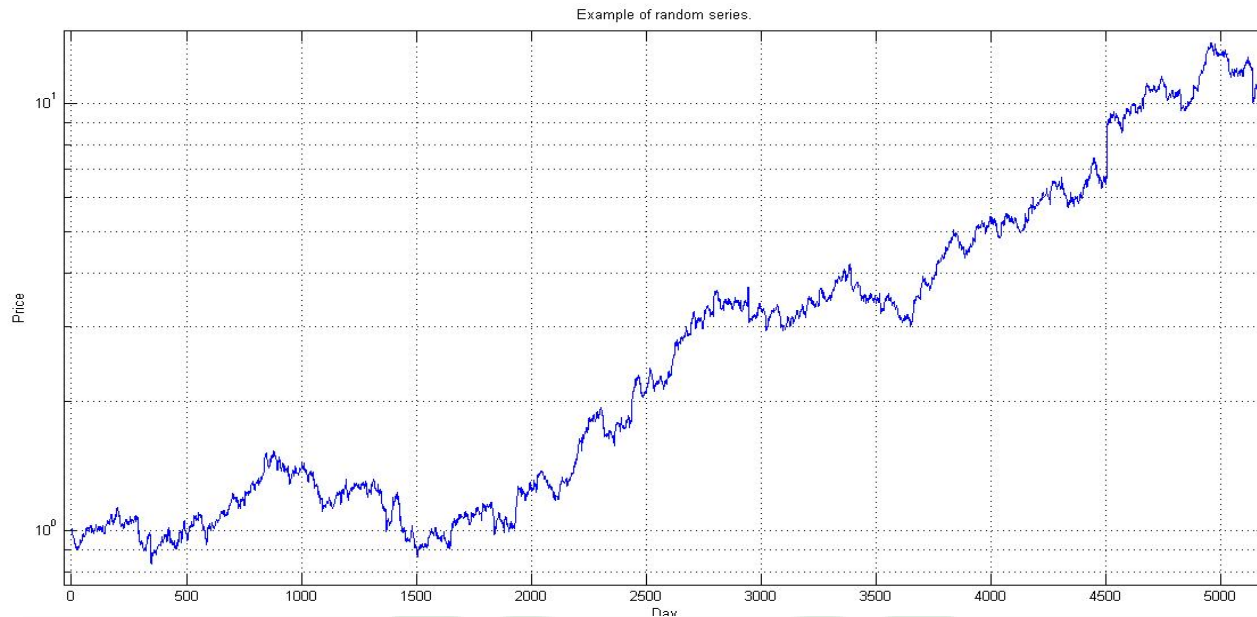
- Las tendencias y estacionalidades pueden darse simultáneamente.



Series Temporales Paramétricas

d) Variaciones aleatorias.

- Los movimientos irregulares (al azar) representan todos los tipos de movimientos de una serie de tiempo que no sea tendencia, variaciones estacionales y fluctuaciones cíclicas.



Series Temporales Paramétricas

➤ El enfoque de descomposición de series de tiempo, supone que la serie $Y(1), \dots, Y(n)$ puede ser expresada como suma o producto de sus cuatro componentes:

- ✓ Tendencia,
- ✓ Componente estacional,
- ✓ Ciclo
- ✓ Término de error aleatorio.



Series Temporales Paramétricas

$$Y(t) = T(t) + E(t) + A(t)$$

Modelo aditivo

$$Y(t) = T(t) \cdot E(t) \cdot A(t)$$

Modelo multiplicativo

donde:

T: Tendencia de la serie.

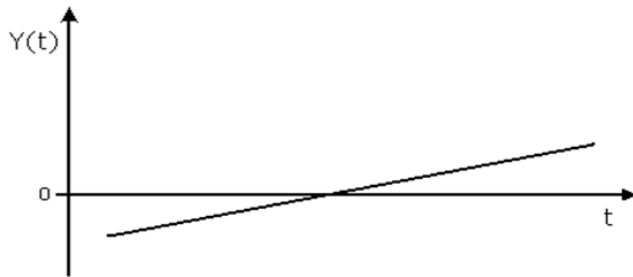
E: Variación Estacional.

A: Variaciones aleatorias.

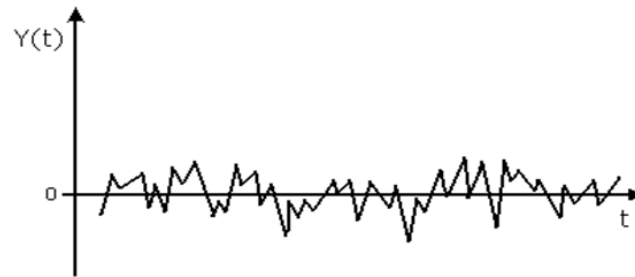


Series Temporales Paramétricas

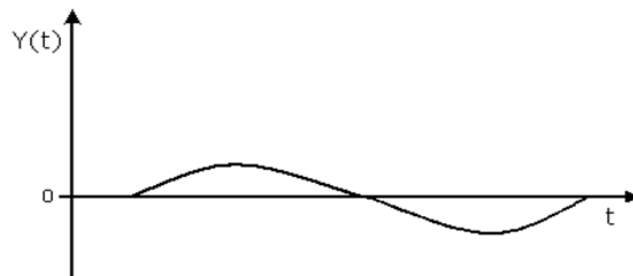
- El gráfico siguiente muestra la serie y sus componentes, para el caso aditivo.



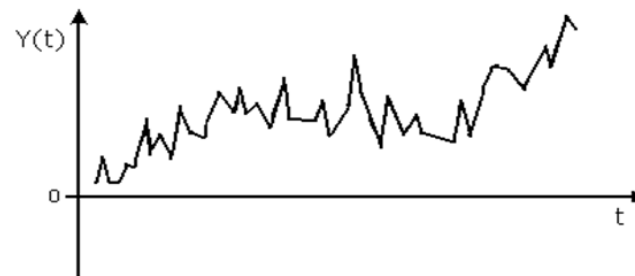
Tendencia



Componente aleatoria



Componente estacional



Serie de tiempo observada



Series Temporales Paramétricas

- Método de Suavización Exponencial Simple (MSES).
- Método de Suavización de Brown.
- Método de Suavización de Holt.
- Método de Suavización de Holt y Winter.



Series Temporales Paramétricas

- Los métodos de suavización exponencial, se basan en suavizar las oscilaciones la serie de tiempo a través de obtener **promedios ponderados** de los valores anteriores.
- Emplea un proceso recursivo de ponderación, que asigna mayor peso a las observaciones más recientes. Esto quiere decir, que las observaciones más recientes tendrán mayor incidencia en el valor del pronóstico que las observaciones que están más alejadas.

$$F_t = \underbrace{F_{t-1}}_{\text{Pronóstico del período anterior}} + \alpha \underbrace{(A_{t-1} - F_{t-1})}_{\text{Error del pronóstico del período anterior}}$$

Pronóstico del
período anterior

Error del pronóstico
del período anterior



Series Temporales Paramétricas

- Recibe el calificativo de exponencial, puesto que la ponderación o pesos de las observaciones decrece exponencialmente a medida que nos alejamos del momento actual t .
- Son métodos para hacer **pronósticos a corto plazo y fáciles de aplicar.**



SERIES TEMPORALES PARAMÉTRICAS I : SUAVIZACIÓN EXPONENCIAL SIMPLE

Ecuación de suavización: $S_t = \alpha Y_t + (1 - \alpha) S_{t-1} \quad t = 1, 2, \dots, n$

S_t = Valor suavizado para el periodo t

α = Constante de suavización ($0 < \alpha < 1$)

Y_t = Valor observado en el periodo t

Se tiene: $S_1 = \alpha Y_1 + (1 - \alpha) S_0$

$$S_2 = \alpha Y_2 + (1 - \alpha) S_1$$

$$S_3 = \alpha Y_3 + (1 - \alpha) S_2$$

- Entonces, para un α grande se esta dando mayor importancia a las observaciones actuales, mientras un valor pequeño damos mucha importancia a las observaciones pasadas.



SERIES TEMPORALES PARAMÉTRICAS I : SUAVIZACIÓN EXPONENCIAL DE BROWN

- El método de suavización Brown o doble (MSED), se usa para pronosticar series de tiempo que presentan una tendencia lineal. El método aplica una primera suavización a los datos originales de la serie y luego una segunda suavización a los datos suavizados de la primera suavización.

Ecuaciones de suavización:

Primera suavización:
$$S'_t = \alpha Y_t + (1 - \alpha) S'_{t-1}$$

Segunda suavización:
$$S''_t = \alpha S'_t + (1 - \alpha) S''_{t-1}$$



SERIES TEMPORALES PARAMÉTRICAS I : SUAVIZACIÓN EXPONENCIAL DE BROWN

Ecuación de pronóstico para m periodos: $\hat{Y}_{t+m} = a_t + b_t m$

Intersección (diferencia de valores suavizados): $a_t = 2 S'_t - S''_t$

Pendiente (factor de ajuste): $b_t = \frac{\alpha}{1-\alpha} (S'_t - S''_t)$

- El MSED usa sólo una constante de suavización y los valores estimados de la tendencia serán muy sensibles a variaciones aleatorias.



SERIES TEMPORALES PARAMÉTRICAS I : SUAVIZACIÓN DE HOLT

- Este método se aplica cuando la serie presenta una tendencia lineal. Utiliza dos constantes diferentes de suavización (suavización biparamétrica), suavizando en forma directa la tendencia y la pendiente de la serie. Este método Holt, proporciona mayor flexibilidad al detectando las proporciones de la tendencia y la pendiente.



SERIES TEMPORALES PARAMÉTRICAS I : SUAVIZACIÓN DE HOLT

Ecuación de suavización:
$$S_t = \alpha Y_t + (1 - \alpha)(S_{t-1} + T_{t-1})$$

- La ecuación proporciona directamente el valor del nivel de la tendencia en el momento t .

Ecuación de la tendencia:
$$T_t = \beta (S_t - S_{t-1}) + (1 - \beta)T_{t-1}$$

- La ecuación permite calcular la tendencia T_t en forma recursiva.
- Para suavizar la estimación de la tendencia se usa una segunda constante de suavización β .



SERIES TEMPORALES PARAMÉTRICAS I : SUAVIZACIÓN DE HOLT

Ecuación de pronóstico para m periodos: $\hat{Y}_{t+m} = S_t + T_t m$

Constantes de suavización. El método de Holt usa dos constantes de suavización, lo cual lo hace desventajoso su determinación. Sin embargo, se puede aplicar el mismo procedimiento que consiste en seleccionar los valores que minimicen la suma de cuadrados del error de suavización.

Valores iniciales. Los valores iniciales para el valor de suavización S_0 y la tendencia T_0 , pueden ser estimados al ajustar por mínimos cuadrados todos los datos a la ecuación de una línea recta. $S_0 = \hat{b}_0$ y $T_0 = \hat{b}_1$



SERIES TEMPORALES PARAMÉTRICAS I : SUAVIZACIÓN DE HOLT Y WINTERS

- Este método se utiliza para obtener pronósticos, cuando la serie presenta una tendencia lineal y variaciones de estacionalidad. Es una extensión del método de Holt, usando tres constantes de suavización. Se usa una ecuación adicional para determinar la estacionalidad.

Ecuación de suavización:
$$S_t = \alpha \frac{Y_t}{E_{t-L}} + (1 - \alpha)(S_{t-1} + T_{t-1})$$

Estimación de la tendencia:
$$T_t = \beta (S_t - S_{t-1}) + (1 - \beta) T_{t-1}$$

Estimación de la estacionalidad:
$$E_t = \gamma \frac{Y_t}{S_t} + (1 - \gamma) S_{t-L}$$

- La estimación de la estacionalidad, está dada por el calculo del índice de estacionalidad $\left(\frac{Y_t}{S_t} \right)$

El valor de L es dado por el periodo estacional.



SERIES TEMPORALES PARAMÉTRICAS I : SUAVIZACIÓN DE HOLT Y WINTERS

Ecuación de pronóstico para m periodos: $\hat{Y}_{t+m} = (S_t - T_t m) E_{t-L+p}$

- La ecuación de pronóstico es válida hasta el periodo ($1 \leq m \leq L$), puesto que se utiliza el factor de estacionalidad de la misma estación del año anterior.

Constantes de suavización. El método de Winter usa tres constantes de suavización, lo cual lo hace desventajoso su determinación. Sin embargo, se puede aplicar el mismo procedimiento que consiste en seleccionar los valores que minimicen la suma de cuadrados del error de suavización.

Valores iniciales. Los valores iniciales necesarios para iniciar el cálculo recursivo son $L+2$ que corresponden a los L índices estacionales del año anterior, la primera observación y al nivel (S_0) y tendencia (T_0).



ENFOQUES PARA EL ANÁLISIS DE SERIES DE TIEMPO

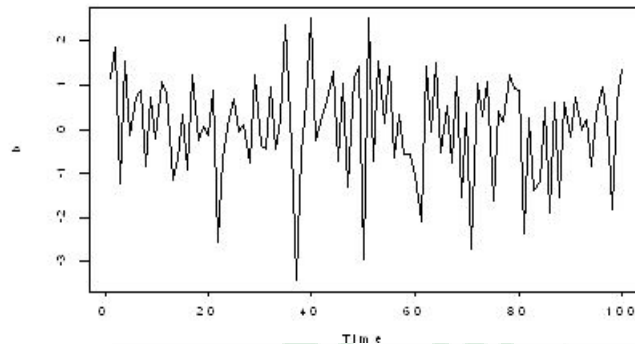
- Los métodos mas utilizados para el análisis de series temporales son:
 - Enfoque de Descomposición
 - **Enfoque de Dominio del Tiempo o Modelos de Box – Jenkins**
 - Enfoque de Dominio de Frecuencias.



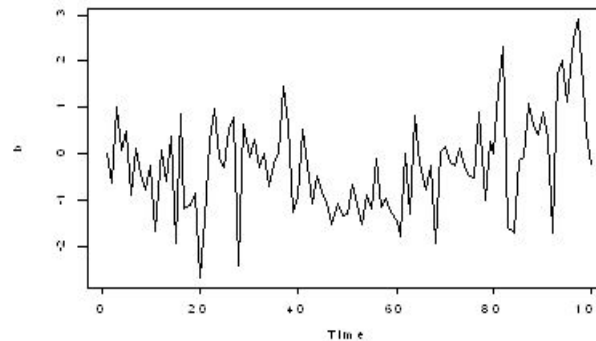
ENFOQUES PARA EL ANÁLISIS DE SERIES DE TIEMPO

ENFOQUE DE DOMINIO DEL TIEMPO

$AR1 = 0.9, MA1 = -0.7$



$AR1 = -0.6, MA1 = 0.3$



MÉTODO DE ENFOQUE DE DOMINIO DEL TIEMPO O MODELOS DE BOX - JENKINS

- Box y Jenkins han desarrollado modelos estadísticos que tienen en cuenta la dependencia existente entre los datos.
- Cada observación en un momento dado es modelada en función de los valores anteriores.
- Se modela a través de ARIMA (Autorregresive Integrate Moving Average).



ENFOQUE DE DOMINIO DEL TIEMPO O MODELOS DE BOX-JENKINS

- Analiza errores recientes de pronósticos para seleccionar el ajuste apropiado para periodos futuros.
- Box-Jenkins es más apropiado para predicciones a largo plazo que para corto plazo.
- Extrae mucha información de la serie de tiempo, más que cualquier otro método.



Proceso: metodología Box-Jenkins

1. Identificación tentativa del modelo



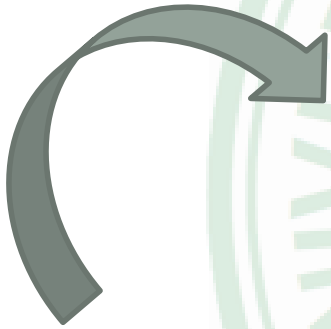
2. Estimación de los parámetros del modelo



3. Evaluación de diagnósticos para comprobar si el modelo es adecuado; mejorar el modelo si es necesario.



4. Generación de Pronósticos



ESQUEMA DE ANÁLISIS : ENFOQUE DE DOMINIO DEL TIEMPO O MODELOS DE BOX - JENKINS



Elección del modelo en el enfoque de Dominio del tiempo

✓ Existen múltiples tipos de modelos en el enfoque de dominio del tiempo:

- Modelos autorregresivos . $AR(p)$
- Modelos de medias móviles. $MA(q)$.
- Modelos Integrados autorregresivos - medias móviles. $ARIMA(p,d,q)$.
- Modelos Integrados autorregresivos - medias móviles estacionales. $SARIMA(P,D,Q)s(p,d,q)$
- Modelos Heterocedásticos. ARCH
- Modelos Espaciales. COARIMA



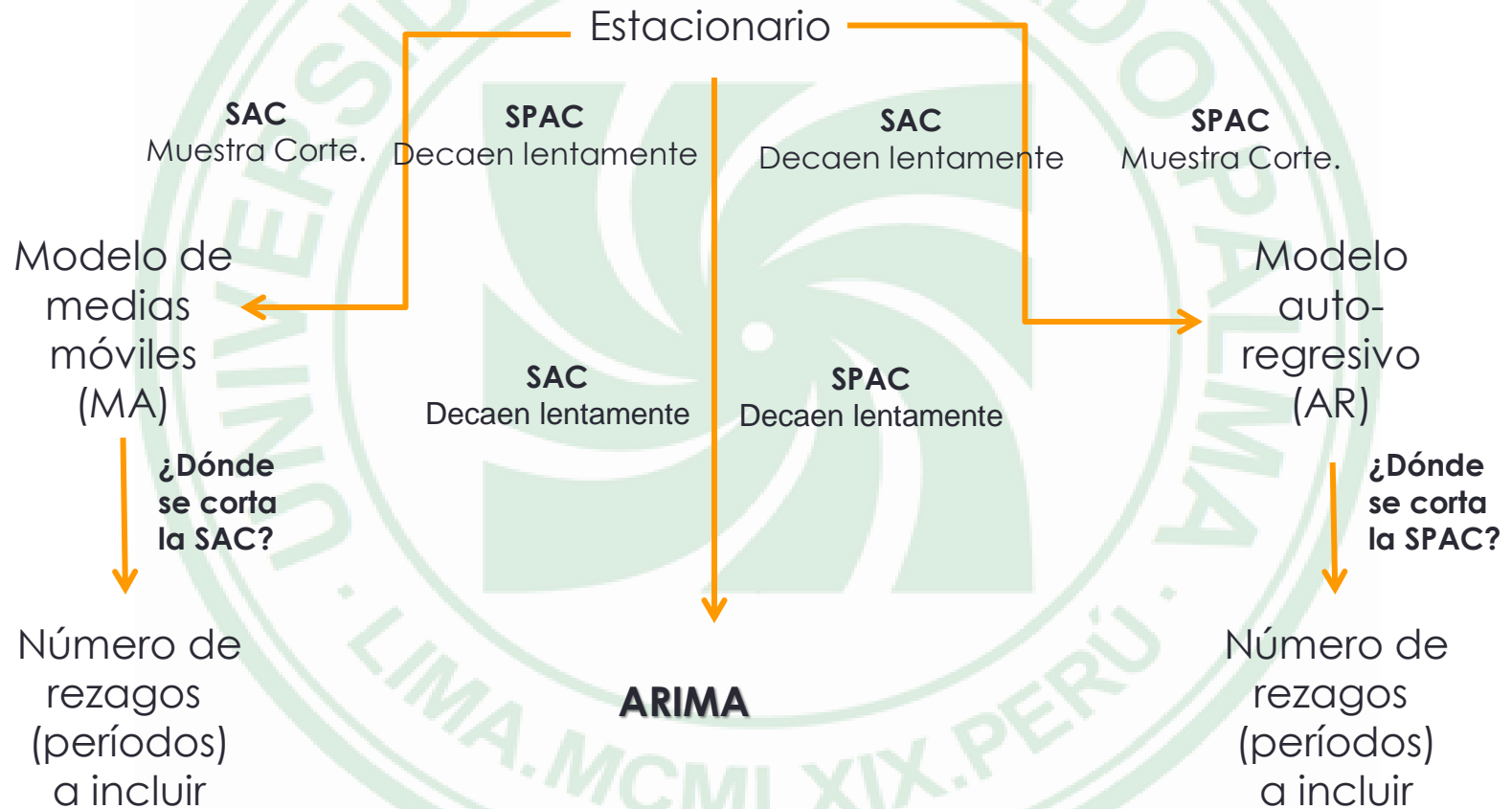
Modelos autorregresivos AR(p).

- ✓ Describe una clase particular de proceso estacionarios, que resultan de imponer una dependencia lineal entre las variables del proceso similar a una ecuación de regresión, el caso mas simple ARIMA (1,0,0) o AR(1).

$$Y_t = \mu + \phi_1 Y_{t-1} + a_t$$

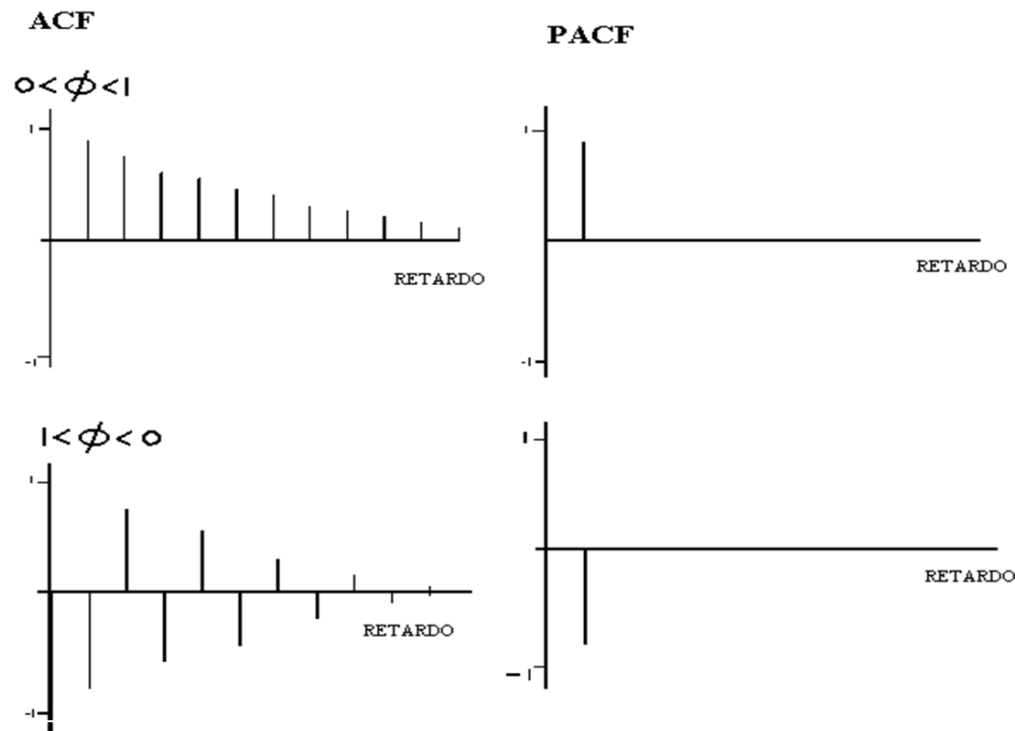


Proceso: Identificación del modelo



AR (1)

- Los procesos AR(1) se reconocen por una ACF infinita y una PACF que desaparece tras el primer retardo. Si los datos tienen media, es necesario especificar un termino constante.



Modelos de medias móviles MA(q)

- ✓ Todo proceso débilmente estacionario de media cero que no contenga componentes determinísticos puede escribirse como una función lineal de variables aleatorias incorrelacionadas a_t , tales como : ARIMA (0,0,1) o MA (1).

$$Y_t = a_t + \Psi a_{t-1}$$



MA(1)

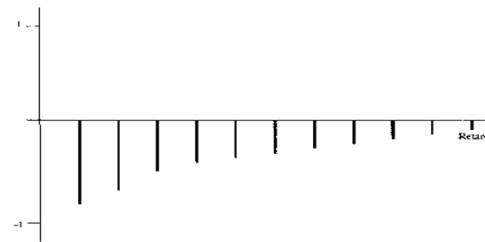
- Los procesos MA(1) se reconocen por una PACF infinita y una ACF que desaparece tras el primer retardo. Si los datos tienen media, es necesario especificar un término constante.

ACF

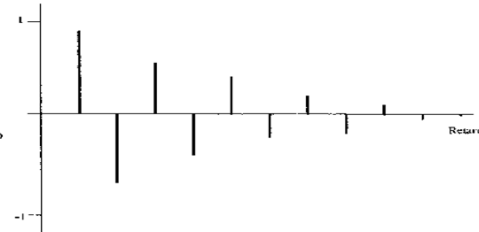
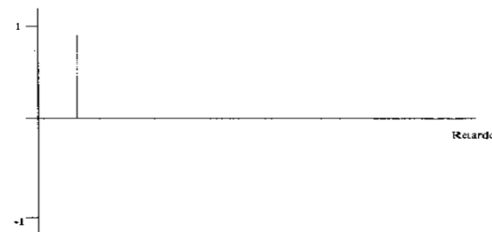
$$0 < \theta < 1$$



PACF



$$-1 < \theta < 0$$



Procesos No Estacionarios: ARIMA(p,d,q)

- Es un modelo que permite describir un valor como una función lineal de los datos rezagados en periodos de tiempo y errores debidos al azar.
- Se dice que una serie sigue un proceso ARIMA(p,d,q) cuando después de aplicarle al operador diferencias d veces sigue un proceso ARMA(p,q) estacionario e invertible.

$$\phi_p(L) \Delta^d x_t = \theta_q(L) \varepsilon_t$$

$$\Delta^d x_t = z_t$$

$$\phi_p(L) z_t = \theta_q(L) \varepsilon_t.$$

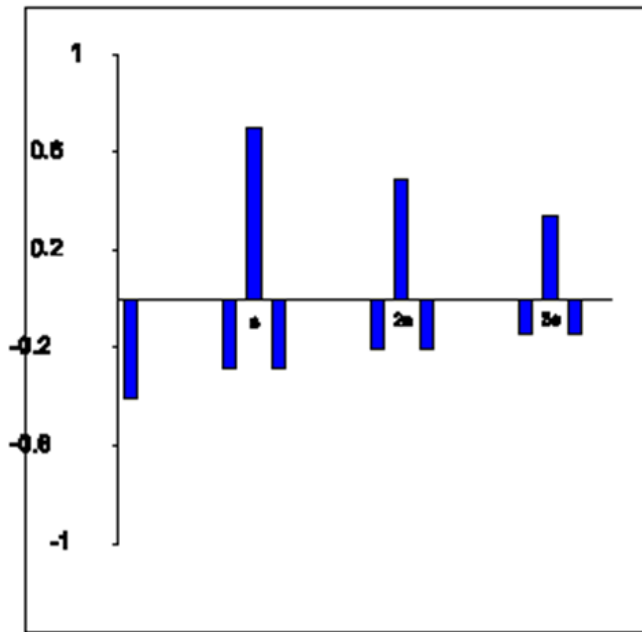


PROCESOS ESTACIONALES: SARIMA(P,D,Q)s(p,d,q)

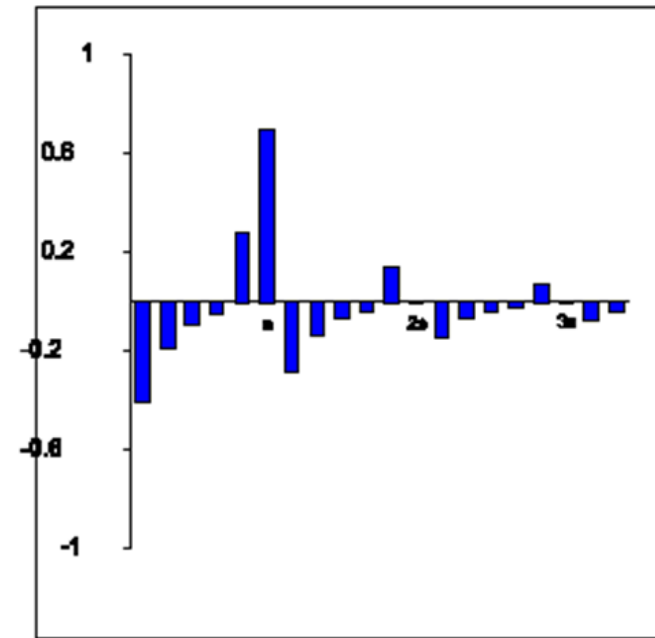
- Si tenemos datos con información de varias ocasiones durante un año, podemos observar estacionalidad. (Ejemplos: temperaturas, vacaciones, movimientos turísticos, etc).
- El número de estaciones durante el año llamamos s . Por ejemplo, $S=12$ para datos mensuales, o 4 para trimestrales. Se puede generalizar los procesos explicados anteriormente para captar estacionalidad.



PROCESOS ESTACIONALES: SARIMA(P,D,Q)s(p,d,q)



Gràfic 31. FAS teòrica d'un
 $MA(1)AR(1)_6$



Gràfic 32. FAP teòrica d'un
 $MA(1)AR(1)_6$

MODELADO BAJO EL ENFOQUE DE DOMINIO DEL TIEMPO O MODELOS DE BOX-JENKINS.

Se postula una clase
de modelos ARIMA

Identificación del modelo
tentativo: p, d, q, P, D, Q

Estimación de
Parámetros del modelo
tentativo

Verificación de
diagnóstico ¿Es adecuado
el modelo?

Uso del modelo con
fines de: Control,
Predicción.



Modelos de series de tiempo

Método de proyección	Cantidad de datos históricos	Patrón de los datos	Horizonte de proyección	Tiempo de preparación	Antecedentes del personal
Ajuste exponencial simple	5 a 10 observaciones para fijar la ponderación	Los datos deben ser estacionarios	Corto	Corto	Poca sofisticación
Ajuste exponencial de Holt	10 a 15 observaciones para fijar la ponderación	Tendencias pero no estacionalidad	Corto a mediano	Corto	Ligera sofisticación
Ajuste exponencial de Winter	Por lo menos 4 ò 5 observaciones por trimestre	Tendencias y estacionalidad	Corto a mediano	Corto	Sofisticación moderada
Modelos de la tendencia de regresión	10 a 20 observaciones para la estacionalidad, por lo menos 5 por trimestre	Tendencias y estacionalidad	Corto a mediano	Corto	Sofisticación moderada
Modelos de regresión causal	10 observaciones por variable independiente	Puede manejar patrones complejos	Corto , mediano o largo	Largo tiempo para el desarrollo ,	Sofisticación considerable
Descomposición de las series de tiempo	Suficiente para ver 2 picos y simas	Maneja patrones cíclicos y estacionales puede identificar los puntos críticos	Corto a mediano	Corto tiempo para la moderación	Poca sofisticación
Box Jenkins	50 o mas observaciones	Deben ser estacionarios o ser transformados en estacionarios	Corto , mediano o largo	Largo	Alta sofisticación



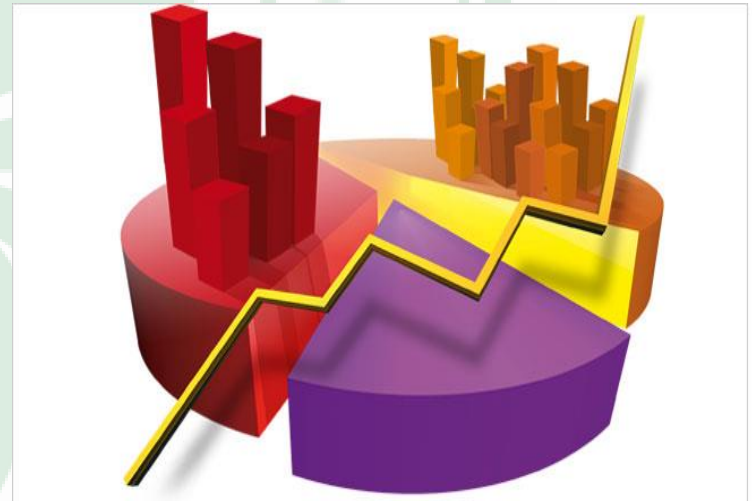
Series temporales paramétricas : Aplicaciones

- ✓ **Economía**: Precios de un artículo, tasas de desempleo, tasa de inflación, índice de precios, precio del dólar, precio del cobre, precios de acciones, ingreso nacional bruto, etc.
- ✓ **Meteorología**: Cantidad de agua caída, temperatura máxima diaria, Velocidad del viento (energía eólica), energía solar, etc.
- ✓ **Geofísica**: Series sismológicas.



Series temporales paramétricas : Aplicaciones

- ✓ **Química:** Viscosidad de un proceso, temperatura de un proceso.
- ✓ **Demografía:** Tasas de natalidad, tasas de mortalidad.
- ✓ **Medicina:** Electrocardiograma, electroencefalograma.
- ✓ **Marketing:** Series de demanda, gastos, utilidades, ventas, ofertas.





¡Gracias!



Comunidad Data Science Perú



Comunidad Data Science Perú

PROGRAMA DE ESPECIALIZACIÓN EN DATA SCIENCE NIVEL I