UNIVERSIDAD DE LOS ANDES
MERIDA VENEZUELA

ESCUELA POLITÉCNICA NACIONAL

# Modelos de Predicción: regresión, series temporales, etc.

**Jose Aguilar**

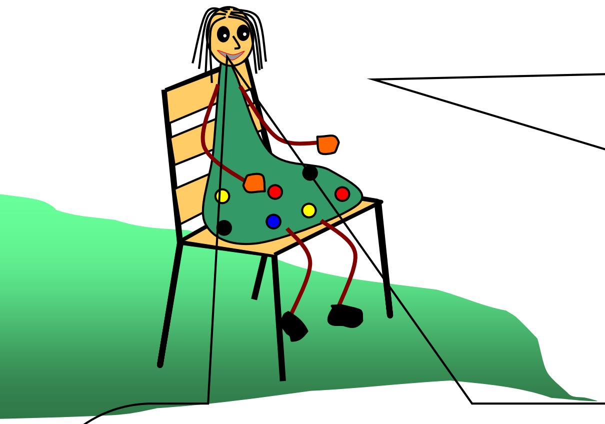
# INTRODUCCION

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, prever o prevenir.

# Pregunta

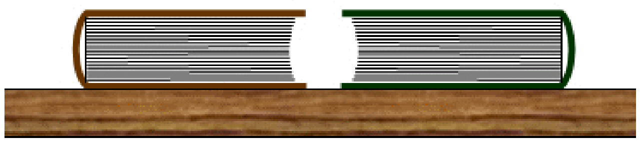


**Se dispone de dos libros con muchas hojas, por ejemplo, dos guías de teléfono.**

Se ponen frente a frente los dos libros y luego, con mucha paciencia, se intercalan las hojas de ellos.

Una vez que se intercalaron todas las hojas, se intenta separarlos tirándolas desde sus respectivos lomos.

**¿Se pueden separar con facilidad?**



**plantee una predicción al respecto.**

Aplicando fuerzas en los lomos respectivos de los libros,

**¿será fácil o difícil separar los libros una vez que tienen mezcladas las hojas?**

# Predicción

- Predice un valor de una variable dada, sobre la base de los valores de otras variables, suponiendo un modelo lineal o no lineal de dependencia.

- **Ejemplos:**
  - Predecir las ventas de nuevos productos basados en gastos de publicidad.
  - Predecir la velocidad del viento como una función de la temperatura, humedad, presión de aire, etc.
  - Predecir comportamiento en el tiempo de los índices bursátiles (series de tiempo).

# Modelos de Predicción

**Las predicciones ayudan a la toma de decisiones en una gran variedad de areas.**

- **Planificación y Control de Operaciones:** Las empresas usan predicciones para decidir que producir, cuando y donde.
- **Mercadeo:** Decisiones de precios, de gastos en publicidad, …dependen fuertemente de las previsiones que se tengan sobre como van a responder las ventas a los diferentes esquemas de marketing.
- **Economía:** Predicciones de las variables macro-económicas claves como el PNB, Paro, Consumo, Inversión, Tipos de Interés, etc… son usadas por el gobierno para fijar su política monetaria y fiscal.
- **Financiera**: actores de los mercados financieros tienen un gran interés en la predicción de los rendimientos de activos financieros (acciones, tipos de interés, tipos de cambio, etc…).
- **Demografía:** La predicción de la población es crucial para planificar el gasto publico en sanidad, infraestructuras, educación, etc.

# Modelos de Predicción

Piensa en una variable que quieras predecir. ***Que necesitas?***

- **Objeto a predecir**: Una serie temporal, un suceso, …etc.
- **Formato de la Predicción**: Puntual, Intervalo, Densidad, …etc.
- **Horizonte de la predicción**: Corto, Medio o Largo Plazo
- **Conjunto de Información**: Univariante o Multivariante
- **Metodos y Complejidad**: Modelos, …etc.

# Modelos de Predicción

Hay muchas formas de hacer predicciones; pero todas ellas tienen en común los siguientes ingredientes:

1. ***que hay ciertas regularidades que captar***
2. ***que tales regularidades son informativas sobre el futuro***
3. ***están encapsuladas en el método seleccionado para predecir***
4. ***normalmente se excluyen las no-regularidades***

Los principales métodos son:

- **Adivinación**
- **Extrapolación**
- **Encuestas**
- **Modelos de Series Temporales**

# Evaluación de Prediciones

Al menos hay tres fuentes de error

- **Incertidumbre en la Especificación:** Todos los modelos están equivocados!!!! (algunos mas que otros)
- **Incertidumbre en la Innovación**: Innovaciones futuras son desconocidas cuando se hace la predicción.
- **Incertidumbre en los Parámetros**: Los coeficientes que usamos para producir las predicciones son *estimaciones,* y por lo tanto están sujetas a variabilidad muestral.

Diferente medidas de errores de predicción

- **Error de Especificación**
- **Error de Aproximación**
- **Error de Estimación**

# Evaluación de Prediciones

**Las medidas mas comunes de la precisión de la predicción son:**

**Error cuadrático medio:**

$$MSE = \frac{1}{T}\sum_{t=1}^{T} e_{t+h,t}^{2}$$

**Raíz cuadrada del MSE**

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T} e_{t+h,t}^{2}}$$

**Error absoluto medio**

$$MAE = \frac{1}{T}\sum_{t=1}^{T} |e_{t+h,t}|$$

donde $\boldsymbol{e_{t+h,t}=y_{t+h}-y_{t+h,t}}$ son los errores de predicción.

# Modelos de Predicción

**El modelo de regresión es un modelo explícitamente multi-variable, en donde la variable a explicar se explica y se predice en base a su propia historia pasada y la historia pasada de otras variables relacionadas.**

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

$$\varepsilon_t \text{ es } WN(0, \sigma^2)$$

# *Predicción*

## Evaluación de la capacidad de predecir

- Dividir la muestra en dos partes; una para estimación del modelo y una para evaluar la capacidad de predecir.
- Estimar el modelo
- Calcular la predicción para los periodos no usadas.
- Compara la predicción con valores reales (error del pronóstico)

# Regresión

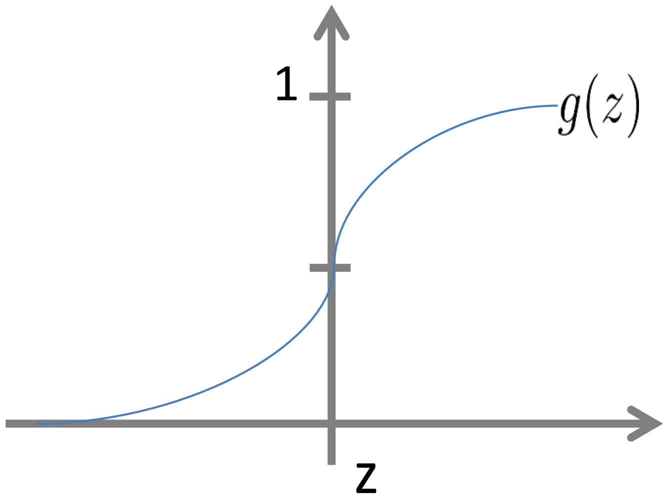
Regresión Lógica : $0 \leq h_\theta(x) \leq 1$

**Cercano a clasificación**

# Regresión Lógica

$$h_\theta(x) = g(\theta^T x)$$

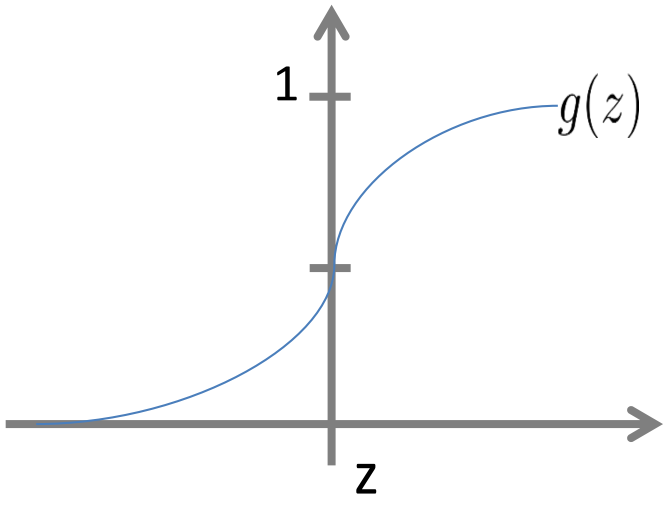
$$g(z) = \frac{1}{1+e^{-z}}$$



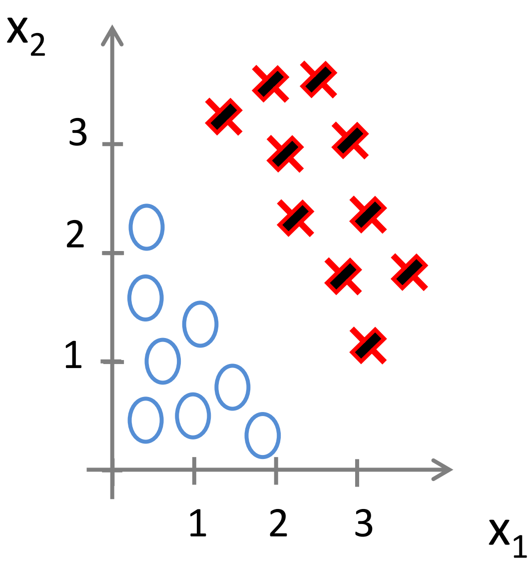
# Regresión Lógica: Barrera de decisión

predice "y=1" si $h_\theta(x) \geq 0.5$

predice "y=0" si $h_\theta(x) < 0.5$

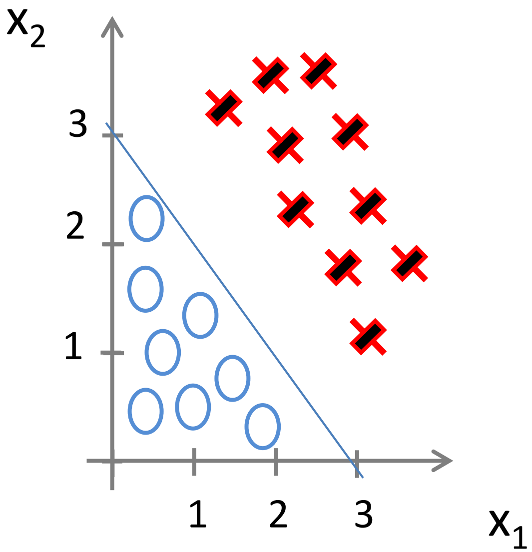


# Regresión Lógica: Barrera de decisión




$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

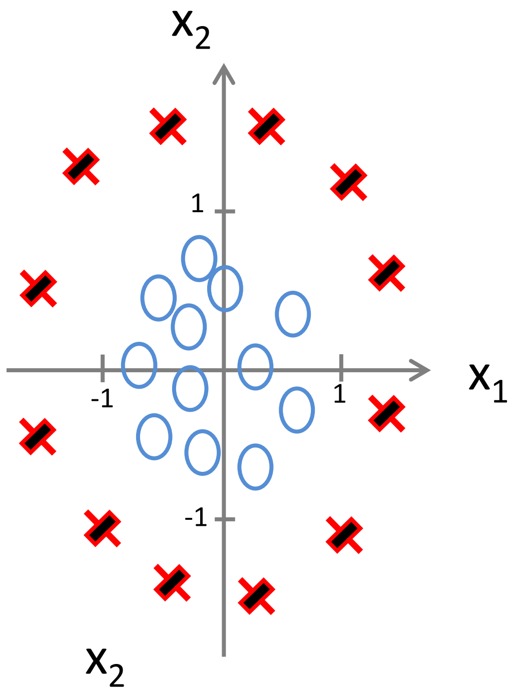
# Regresión Lógica: Barrera de decisión




$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2)$$

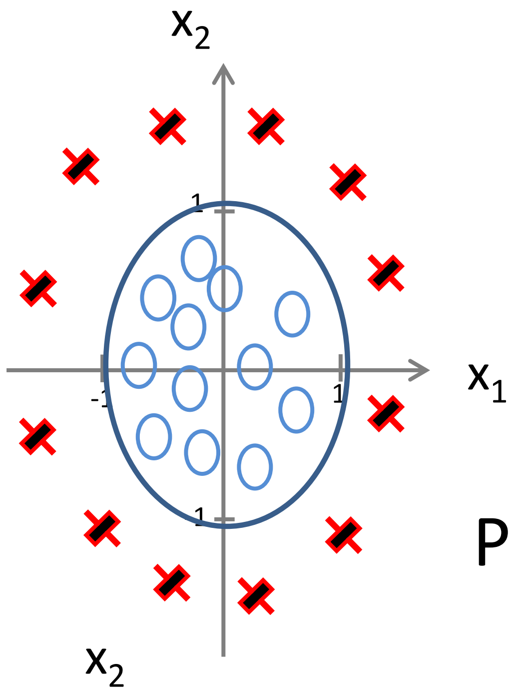
Predice “y=1“ si $-3 + x_1 + x_2 \geq 0$

# Regresión Lógica: Barrera de decisión




$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

# Regresión Lógica: Barrera de decisión




$$h_\theta(x) = g(\theta_0 + \theta_1 x_1 + \theta_2 x_2 + \theta_3 x_1^2 + \theta_4 x_2^2)$$

Predice " y=1" si $-1 + x_1^2 + x_2^2 \geq 0$

# Regresión Lógica: Entrenamiento y Función de costos

Conjunto de entrenamiento :

$$\{(x^{(1)}, y^{(1)}), (x^{(2)}, y^{(2)}), \cdots, (x^{(m)}, y^{(m)})\}$$

n ejemplos

$$x \in \begin{bmatrix} x_0 \\ x_1 \\ \cdots \\ x_n \end{bmatrix} \qquad y \in \{0, 1\}$$

$$h_\theta(x) = \frac{1}{1 + e^{-\theta^T x}}$$

**Como escoger el parámetro $\theta$ ?**

# Regresión Lógica: Función de costos

$$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)})$$

$$\mathrm{Cost}(h_\theta(x), y) = \begin{cases} -\log(h_\theta(x)) & \text{if } y = 1 \\ -\log(1 - h_\theta(x)) & \text{if } y = 0 \end{cases}$$

$$\min_\theta J(\theta)$$

# Gradiente descendiente

$$J(\theta) = -\frac{1}{m}[\sum_{i=1}^{m} y^{(i)} \log h_\theta(x^{(i)}) + (1-y^{(i)}) \log(1-h_\theta(x^{(i)}))]$$

Queremos $\min_\theta J(\theta)$ :

Repeat $\{$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$ (simultaneamente actualizar $\theta_j$ )

$\}$

# Función de costos

**Regresión lineal:** $$J(\theta) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{2} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

$$\mathrm{Cost}(h_\theta(x^{(i)}), y^{(i)}) = \frac{1}{2} \left(h_\theta(x^{(i)}) - y^{(i)}\right)^2$$

# Problema de Optimización!!!

Dado $\theta$ , queremos calcular

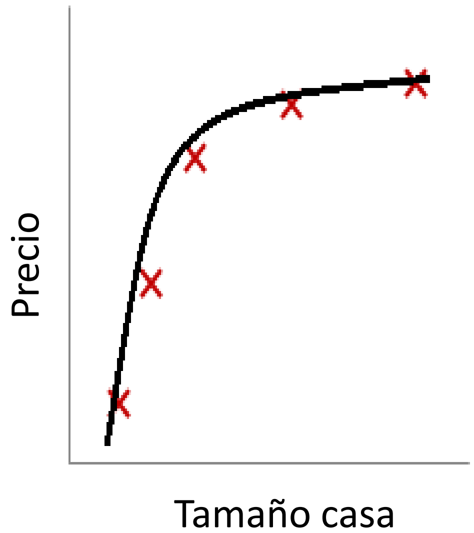
- $J(\theta)$
- $\frac{\partial}{\partial \theta_j} J(\theta)$

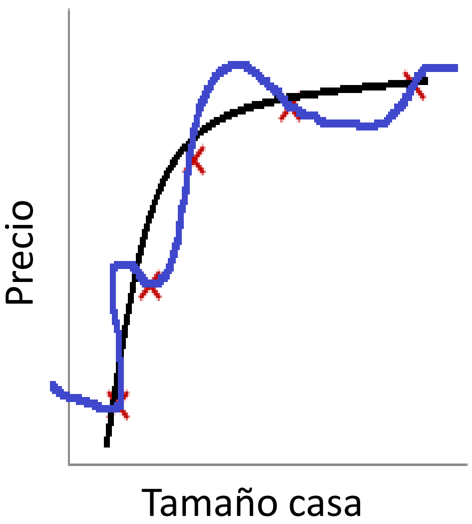
(for j=0, …, n) $\{$

$$\theta_j := \theta_j - \alpha \frac{\partial}{\partial \theta_j} J(\theta)$$

$\}$

# Función de costo




$$\theta_0 + \theta_1 x + \theta_2 x^2$$

$$\theta_0 + \theta_1 x + \theta_2 x^2 + \theta_3 x^3 + \theta_4 x^4$$

**Penalizar** $\theta_3 \; \theta_4$

$$\min_{\theta} \frac{1}{2m} \sum_{i=1}^{m} (h_\theta(x^{(i)}) - y^{(i)})^2$$

# Patrones Secuenciales

- Descubrir patrones en los cuales la presencia de un conjunto de ítems es seguido por otro ítem en orden temporal.

- Ejemplo: Encontrar y predecir el comportamiento de los visitantes de un sitio Web con respecto al tiempo.

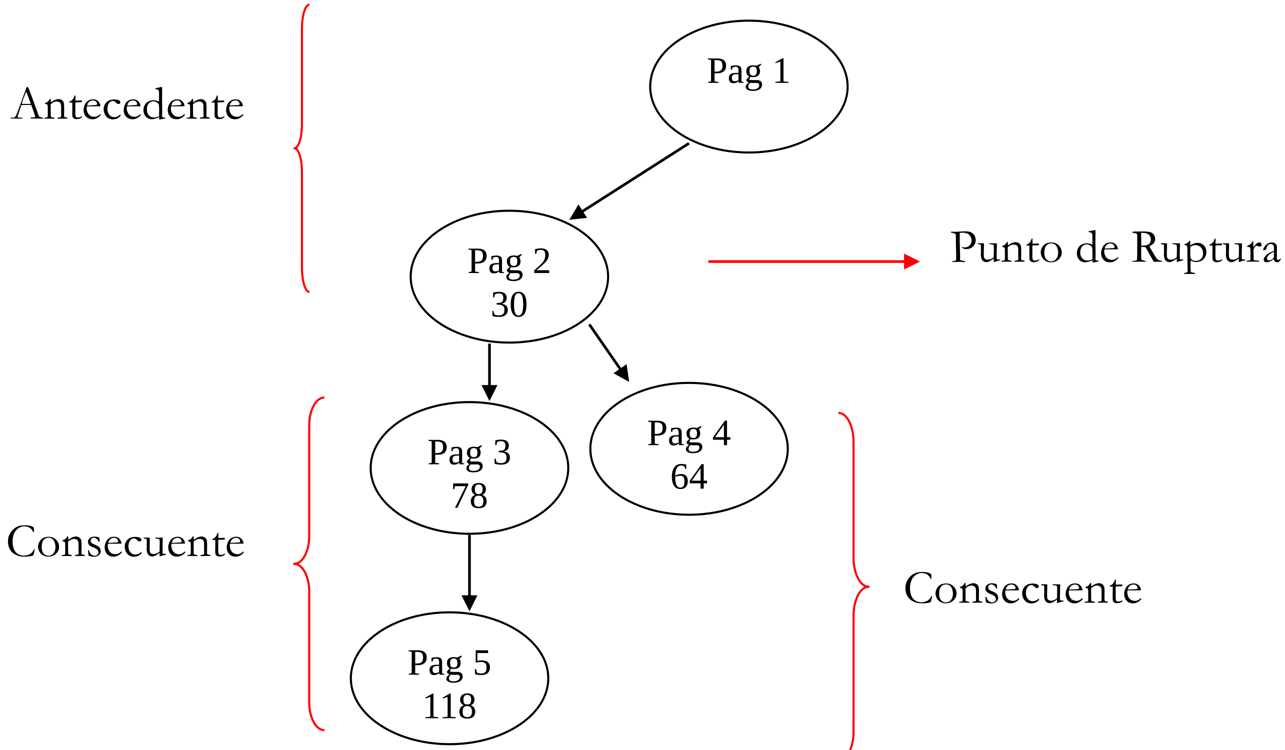
$$[x1 \rightarrow x2 \rightarrow x3] \rightarrow [y1 \rightarrow y2] \text{ en t días}$$

[/public/team.jsp ->
/public/findUsers.jsp->
/private/mycourses/website/folders/assignment/assignment_view.jsp->
/public/portalDocument.js

**en 2 días**

# Patrones Secuenciales

## Generación FBP-Árbol (Matriz FTM, Lista de Caminos)



# Patrones Secuenciales

## Algoritmo Patrones (FBP-Arbol, soporte, confianza)

La confianza de una *regla de comportamiento-frecuente* se representa como *conf*(PIND → PDEP) y define la probabilidad de recorrer el camino PDEP una vez se ha recorrido el camino PIND.

- Se recorre el árbol desde las hojas al nodo raíz.
- Teniendo en cuenta el soporte de cada camino, las reglas son calculados como sigue.
- Buscar en hojas el punto de ruptura.
  - Si la hoja no es Punto ruptura, ir a hoja anterior.
  - Si la hoja es Punto Ruptura, calcular *conf*.
    - Si *conf* > *confianza*, genera Patrón
    - Si *conf* < *confianza*, podar rama de árbol.

## DEFINICION BÁSICA DE SERIE DE TIEMPO

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 intevalos de tiempo iguales) .

Las observaciones de una serie de tiempo serán denotadas por

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

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

# SERIES TEMPORALES

- La serie puede ser simple o multiple según se disponga de un valor para cada instante del tiempo o varios valores.
- Puede ser contínua o discreta, según la naturaleza de la variable dependiente.
- Tiene que estar perfectamente ordenado.
- Los intervalos de tiempo considerado, han de tener la misma amplitud.
- Los datos han de ser homogéneos, no se puede cambiar de criterios, metodología, etc..

# Ejemplos de series de tiempo

1. Economía: Precios de un articulo, tasas de desempleo, tasa de inflación, índice de precios, precio del dólar, precio del cobre, precios de acciones, ingreso nacional bruto, etc.
2. Meteorología: Cantidad de agua caída, temperatura máxima diaria, Velocidad del viento (energía eólica), energía solar, etc.
3. Geofísica: Series sismológicas.
4. Química: Viscosidad de un proceso, temperatura de un proceso.
5. Demografía: Tasas de natalidad, tasas de mortalidad.

## Ejemplos de series de tiempo

6. Medicina: Electrocardiograma, electroencéfalograma.

7. Marketing: Series de demanda, gastos, utilidades, ventas, ofertas.

8. Telecomunicaciones: Análisis de señales.

9. Transporte: Series de tráfico.

...y muchos otros.

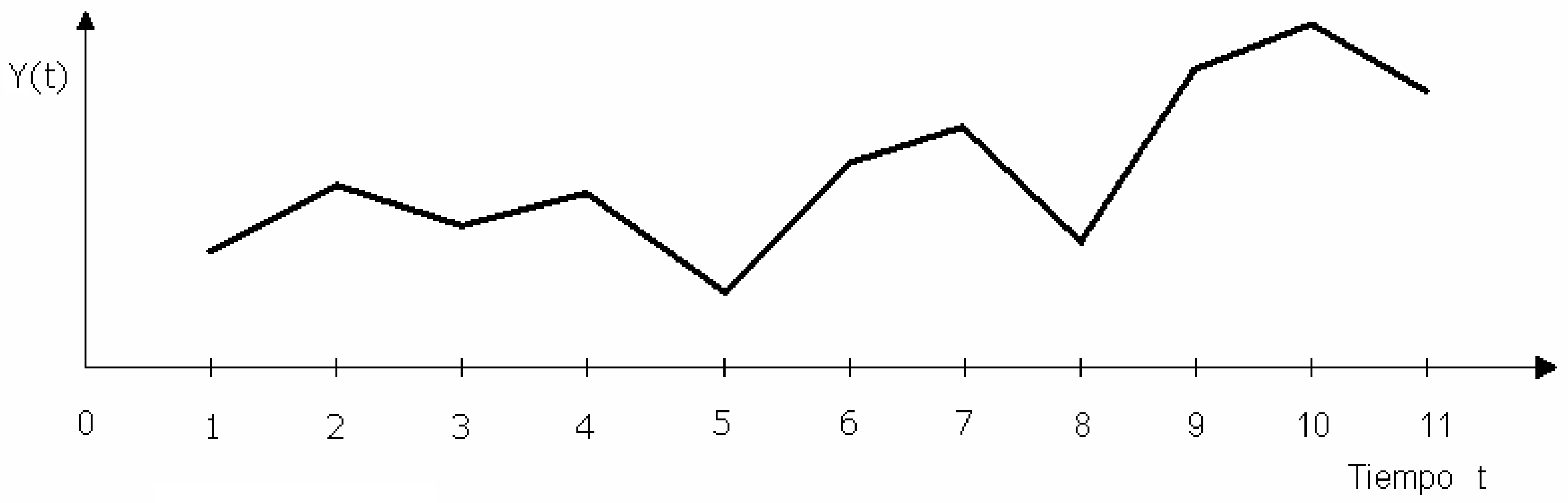
# SERIES TEMPORALES

- **Método que se centran en el dominio del tiempo:** evolución temporal.
  - Análisis Clásico.
  - Metodología Box-Jenkins

- **Método que se centra en el estudio de las frecuencias.**
  - Análisis espectral.

# ANALISIS GRAFICO DE UNA SERIE DE TIEMPO

Por muy simple que parezca, el paso más importante en el análisis de series de tiempo consiste en graficar la serie.




Esto debe hacerse siempre, independiente de cuán simples o complejos sean los procedimientos que se emplean posteriormente.

# Predicción por descomposición de series

## Descomposición de series:

Sobre una serie temporal **Yt** podemos identificar una serie de componentes básicos que se denominan respectivamente como:

•TENDENCIA: **Tt** Movimientos de larga duración que se mantienen durante todo el periodo de observación.

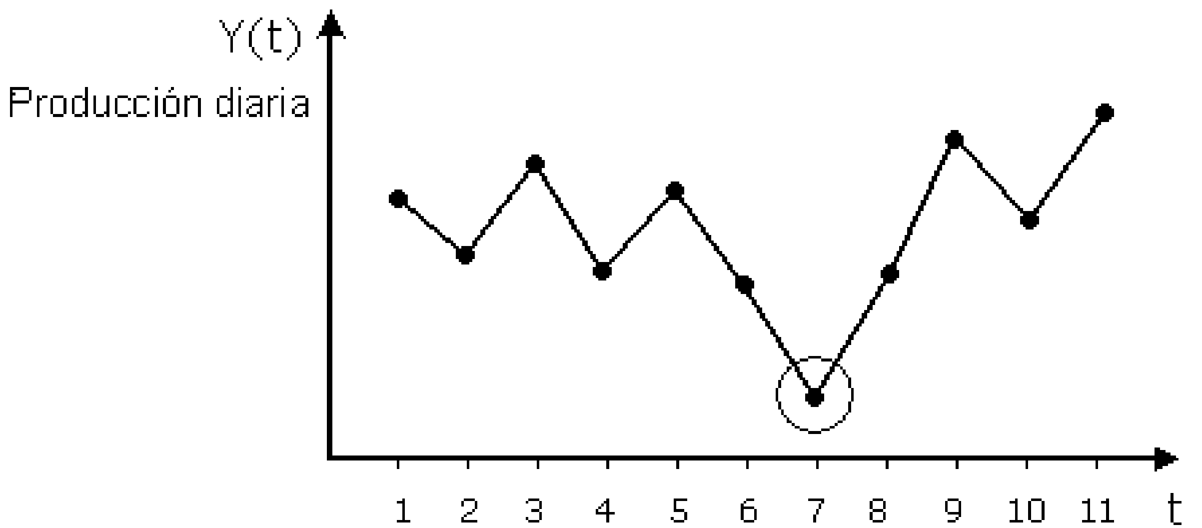
•CICLO: **Ct** Oscilaciones alrededor de la tendencia producidos por períodos alternativos de prosperidad y depresión.

•ESTACIONALIDAD: **St** Movimiento que se produce, dentro de un periodo anual, por motivos no estrictamente económicos (climáticos, sociales,ect.)

•IRREGULARIDAD: **It** Movimientos erráticos generados por causas ajenas al fenómeno económico y no repetidos en el tiempo

## a) <u>Outliers</u>:

Se refiere a puntos de la serie que se escapan de lo normal.




Movimientos que no responden a ningún patrón que son resultado de factores fortuitos o aleatorios.

Puntos de ruptura, interrupciones en la evolución de la serie.

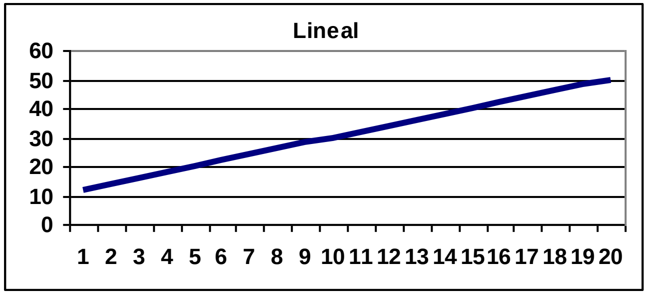
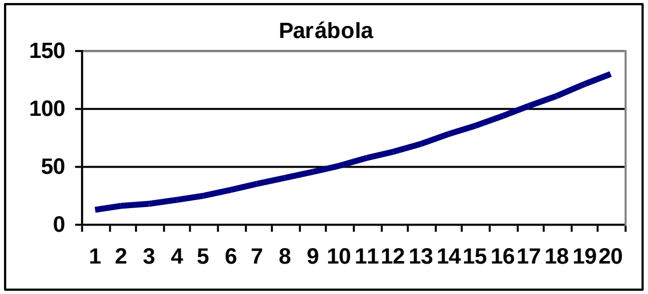
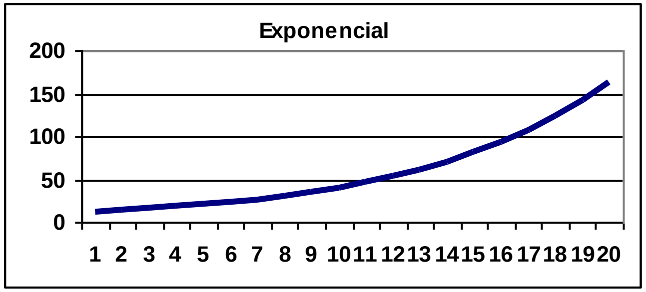
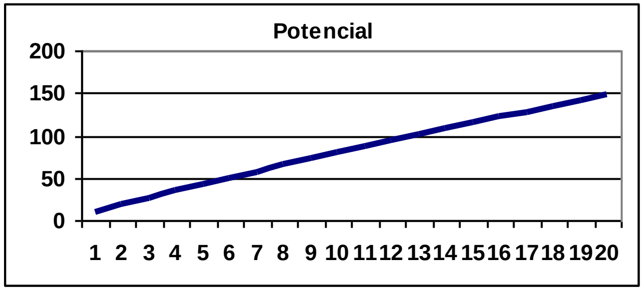
b) **Tendencias**

La tendencia representa el comportamiento predominante de la serie.

# TENDENCIA (T)

- Comportamiento de la serie a l/pl. (+10 años)
- Mientras más larga sea la serie mejor.
- Nos muestra si la serie es estacionaria o evolutiva.
- El comportamiento tendencial puede ser:
  - Lineal
  - Exponencial
  - Parabólica
  - ….

# Predicción por descomposición de series

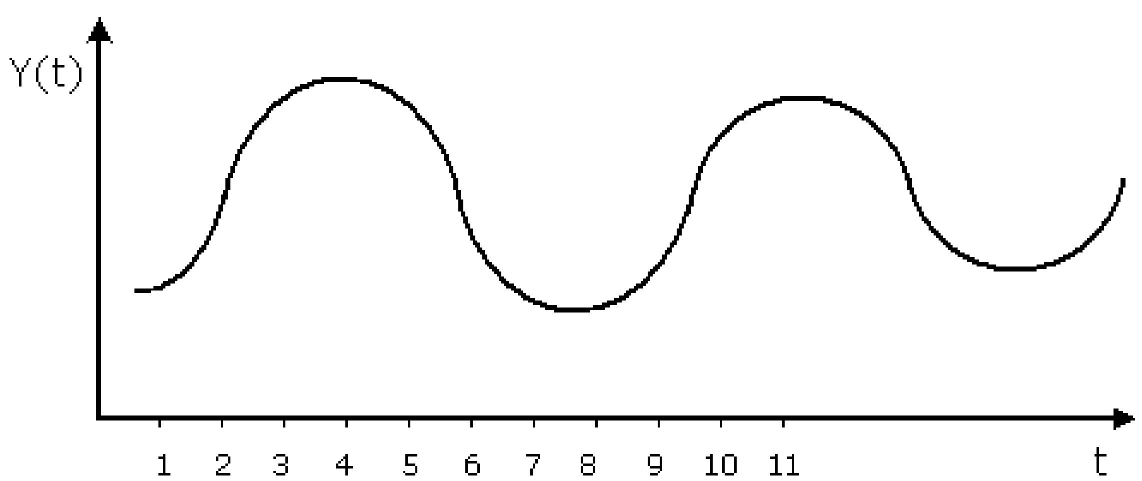
<table>
<tr><td>Lineal</td><td>$$Y_t = a + b * t$$</td><td>
</td></tr>
<tr><td>Parábola</td><td>$$Y_t = a + b * t + c * t^2$$</td><td>
</td></tr>
<tr><td>Exponencial</td><td>$$Y_t = a * b^t$$ $$Ln(Y_t) = c + d * t$$</td><td>
</td></tr>
<tr><td>Potencial</td><td>$$Y_t = a * t^b$$ $$Ln(Y_t) = c + d * Ln(t)$$</td><td>
</td></tr>
</table>

## c) **Variaciones cíclicas o estacionales**

Estas variaciones representa un movimiento *periódico* de la serie de tiempo.

- La **variación estacional** ocurre 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.

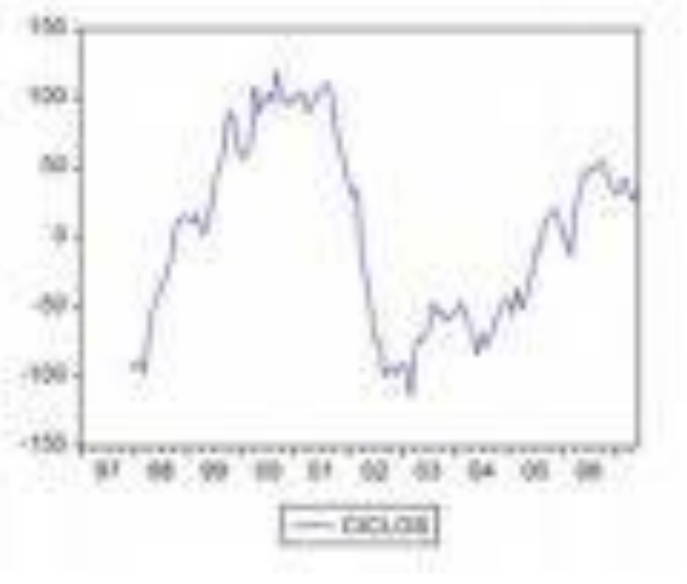
La duración del período puede ser un año, un trimestre, un mes, un día, etc.



# VARIACIONES CÍCLICAS (C)

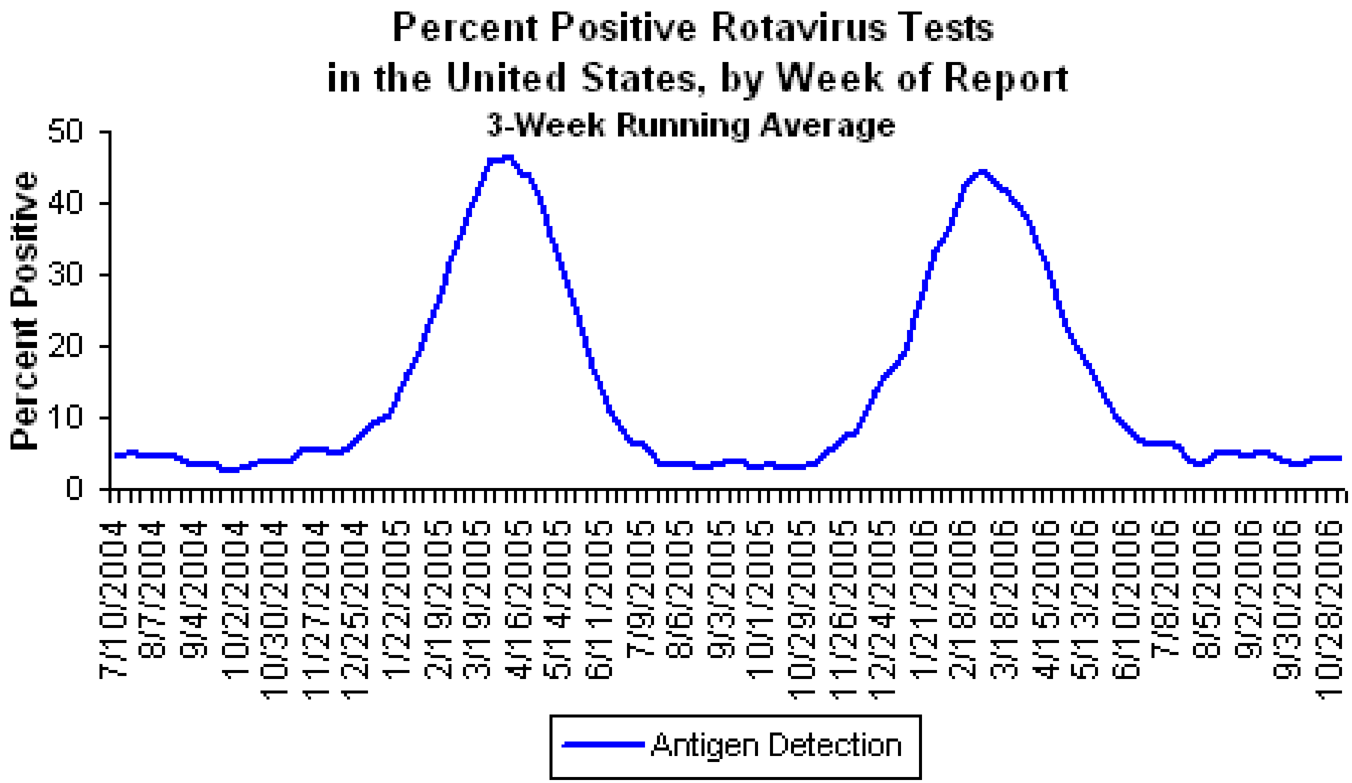
- Son movimientos a medio plazo que corresponden normalmente con las fases expansivas y recesivas de la economía. (Entre 2 y 10 años).

- Son difíciles de detectar, salvo que se tenga un periodo largo de años.

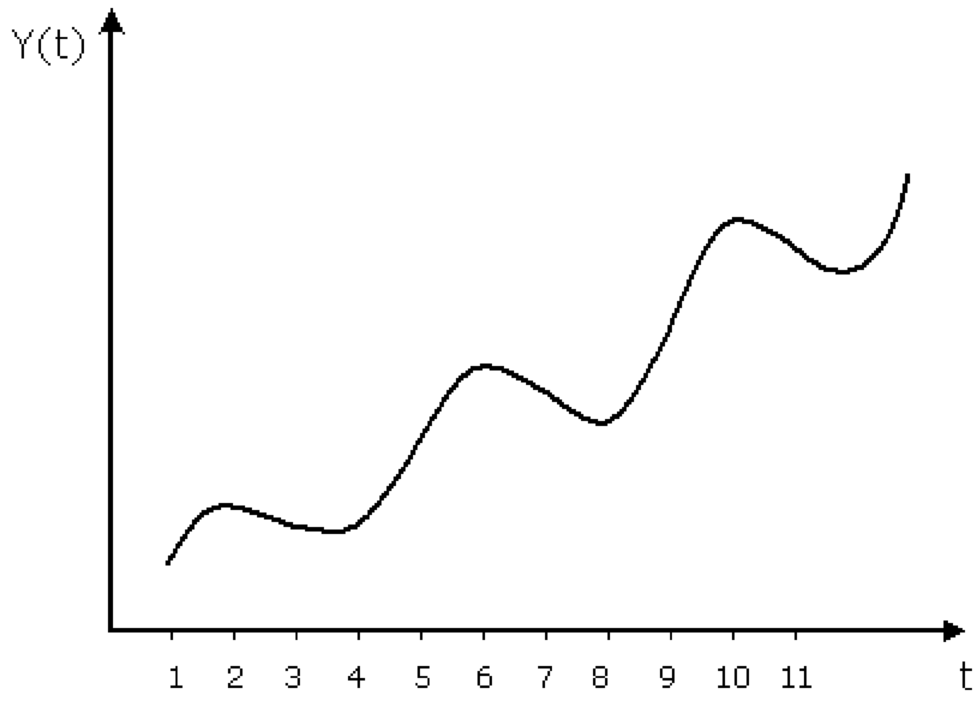


# VARIACIONES ESTACIONALES (VE)

- Movimientos que se repiten de forma periódica, producida normalmente por el clima, producción, tiempo …. Lo usual es que tenga una frecuencia anual.



Las tendencias y estacionalidades pueden darse simultáneamente.



# Modelos Clásicos

Un modelo clásico de series de tiempo, supone que la serie Y(1), ..., Y(n) puede ser expresada como suma o producto de tres componentes:

tendencia,

componente estacional,

término de error aleatorio.

## Modelos Clásicos

1. $Y(t) = T(t)+E(t)+A(t)$ Modelo aditivo

2. $Y(t) = T(t)\ E(t)\ A(t)$ Modelo multiplicativo

donde:

T: Tendencia de la serie.

E: Variación Estacional.

A: Variaciones aleatorias.

# Predicción por descomposición de series

## Descomposición de series:

Podemos plantear diferentes esquemas alternativos de descomposición de una serie temporal:

- ADITIVO:

$$Y_t = T_t + C_t + S_t + I_t$$

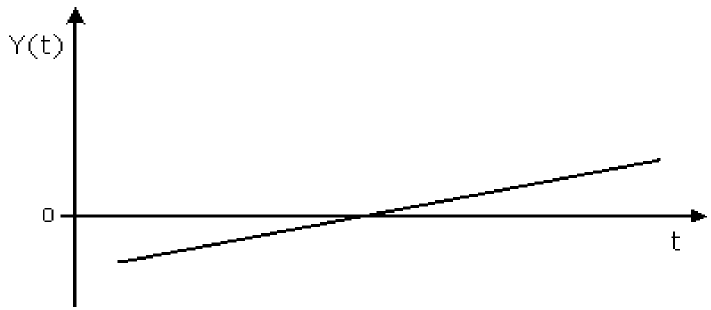
- MULTIPLICATIVO:

- MIXTO:

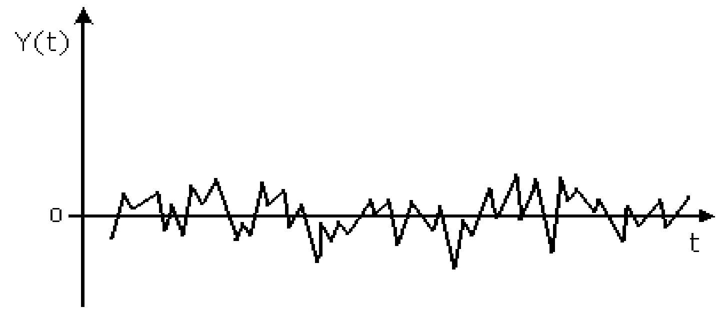
$$Y_t = T_t * C_t * S_t * I_t$$

$$Y_t = T_t * (1 + C_t) * (1 + S_t) + I_t$$

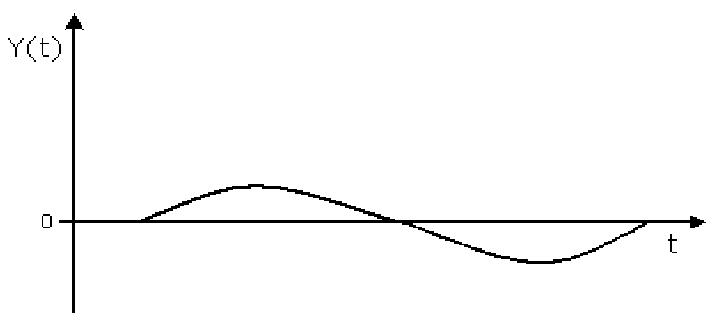
## La serie y sus componentes, para el caso aditivo.



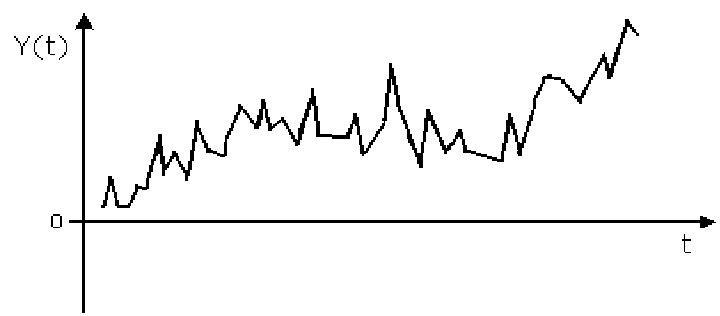
Tendencia



Componente aleatoria



Componente estacional



Serie de tiempo observada

El problema que se presenta es modelar adecuadamente las componentes de la serie.

## ESTIMACIÓN DE LA TENDENCIA

Hay varios métodos para estimar la tendencia T(t), uno de ellos es utilizar un modelo de regresión lineal.

Se pueden utilizar otros tipos de regresiones, como regresión cuadrática, logística, exponencial, entre otros.

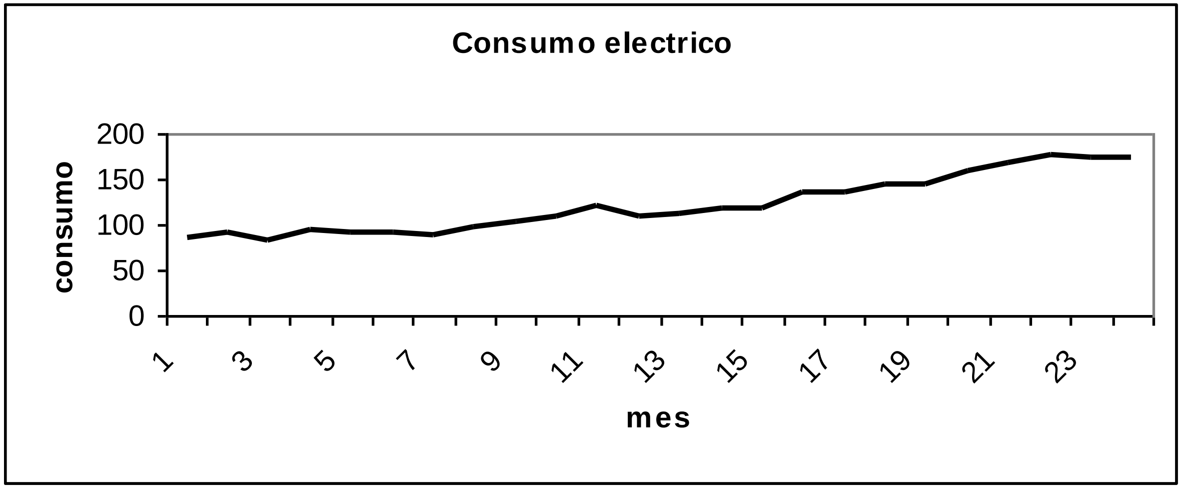
EJEMPLO 1: La tabla presenta parte de los datos de una serie de energía eléctrica. Son 24 datos mensuales referentes a los años 1977 a 1978.

Consumo de Energía Eléctrica

| **t** | **Y(t)** |
|---|---|
| 1 | 84,6 |
| 2 | 89,9 |
| 3 | 81,9 |
| 4 | 95,4 |
| 5 | 91,2 |
| 6 | 89,8 |
| 7 | 89,7 |
| 8 | 97,9 |
| 9 | 103,4 |
| 10 | 107,6 |
| 11 | 120,4 |
| 12 | 109,6 |

| **t** | **Y(t)** |
|---|---|
| 13 | 110,3 |
| 14 | 118,1 |
| 15 | 116,5 |
| 16 | 134,2 |
| 17 | 134,7 |
| 18 | 144,8 |
| 19 | 144,4 |
| 20 | 159,2 |
| 21 | 168,2 |
| 22 | 175,2 |
| 23 | 174,5 |
| 24 | 173,7 |

Gráfico de la serie:




El modelo de tendencia propuesto es un modelo de regresión lineal:

$$Y(t) = \beta_0 + \beta_1 t + A(t)$$

Recurriendo al método de **mínimos cuadrados** se estiman los parámetros y se obtiene

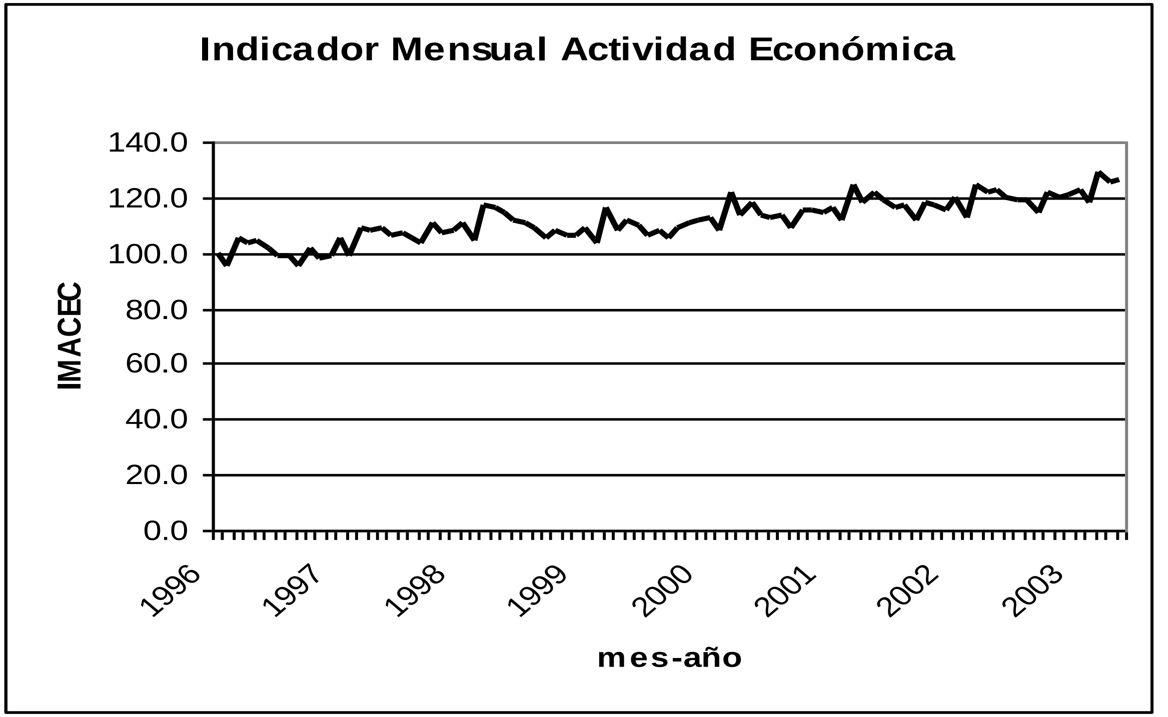
$$T(t) = 68.45 + 4.24 * t$$

## EJEMPLO 2.

Indicador Mensual de Actividad Económica (IMACEC). Base del índice : 1996=100

Corresponde al Indicador Mensual de Actividad Económica (Imacec),

| | 1996 | 1997 | 1998 | 1999 | 2000 | 2001 | 2002 | 2003 |
|---|---|---|---|---|---|---|---|---|
| Enero | 99.6 | 105.0 | 110.8 | 109.2 | 112.6 | 116.4 | 119.7 | 122.6 |
| Febrero | 94.9 | 98.6 | 104.3 | 103.7 | 107.6 | 111.8 | 113.0 | 118.3 |
| Marzo | 105.4 | 109.1 | 117.5 | 116.4 | 121.2 | 124.3 | 124.4 | 128.8 |
| Abril | 103.4 | 108.1 | 116.1 | 108.0 | 113.8 | 118.0 | 122.0 | 125.3 |
| Mayo | 104.2 | 109.2 | 114.4 | 111.2 | 117.9 | 121.7 | 123.0 | 126.1 |
| Junio | 101.3 | 106.5 | 111.9 | 110.0 | 113.1 | 119.1 | 120.1 | |
| Julio | 98.7 | 107.1 | 110.9 | 106.4 | 112.3 | 116.0 | 118.9 | |
| Agosto | 98.7 | 105.6 | 109.0 | 108.1 | 113.4 | 116.9 | 119.1 | |
| Septiembre | 94.8 | 103.8 | 105.4 | 105.7 | 108.6 | 111.4 | 114.6 | |
| Octubre | 102.0 | 110.9 | 107.7 | 109.2 | 115.4 | 118.4 | 121.7 | |
| Noviembre | 98.0 | 106.8 | 106.1 | 110.7 | 114.9 | 117.3 | 119.9 | |
| Diciembre | 99.2 | 108.4 | 106.5 | 111.9 | 114.4 | 115.7 | 120.9 | |


Indicador Mensual Actividad Económica
IMACEC
140.0
120.0
100.0
80.0
60.0
40.0
20.0
0.0
1996
1997
1998
1999
2000
2001
2002
2003
mes-año

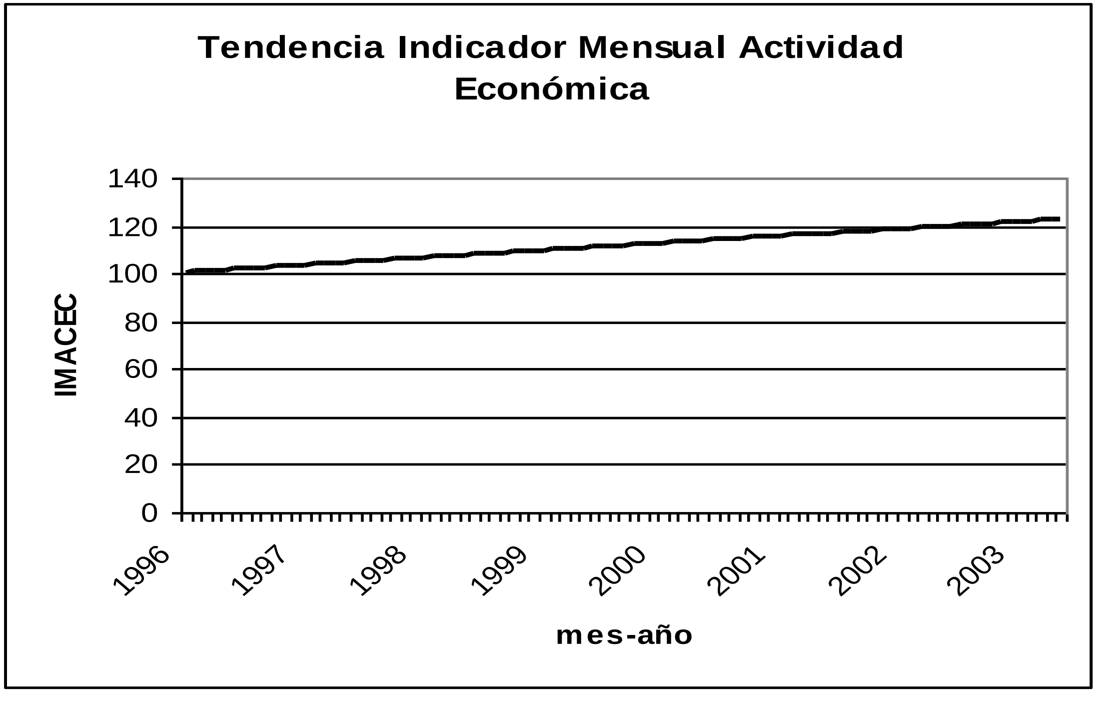
Se estima la tendencia por el método de mínimos cuadrados, de regresión lineal

$$Y(t) = a + bt + A(t)$$

dando el siguiente resultado:

Intercepto **a = 100.3**. Corresponde al valor de partida.

Pendiente **b = 0.253**. Corresponde al aumento medio mensual.




La recta de regresión correspondiente a la tendencia

## ESTIMACION DE LA COMPONENTE ESTACIONAL

Para estimarla, se debe conocer el período, y se deben tener datos de varios períodos consecutivos.

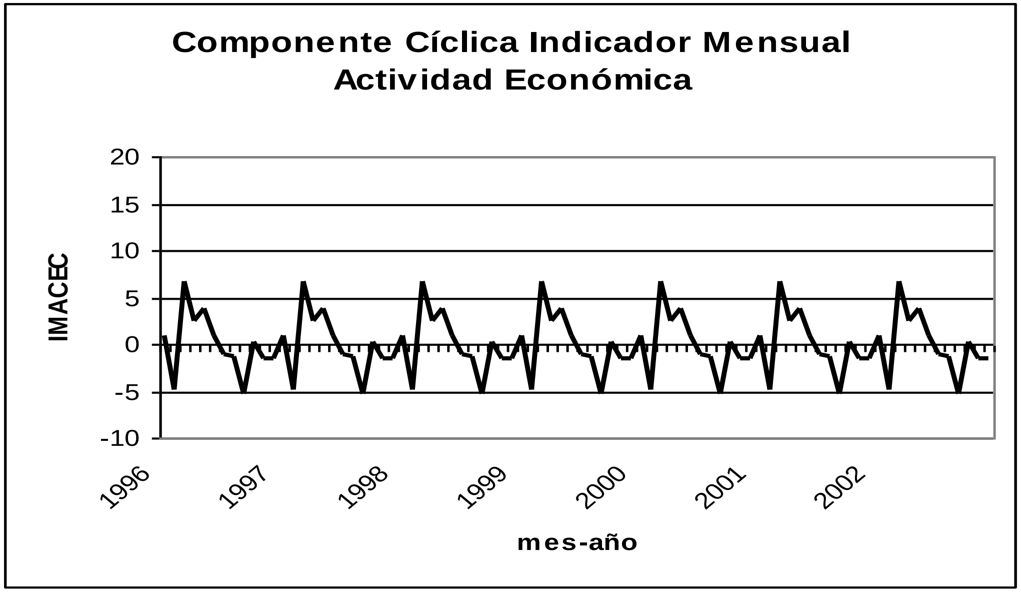
Por ejemplo, datos mensuales, estacionalidad de un año.

El ejemplo siguiente ilustra la forma de obtener la componente estacional.

Asumiremos un modelo clásico aditivo.

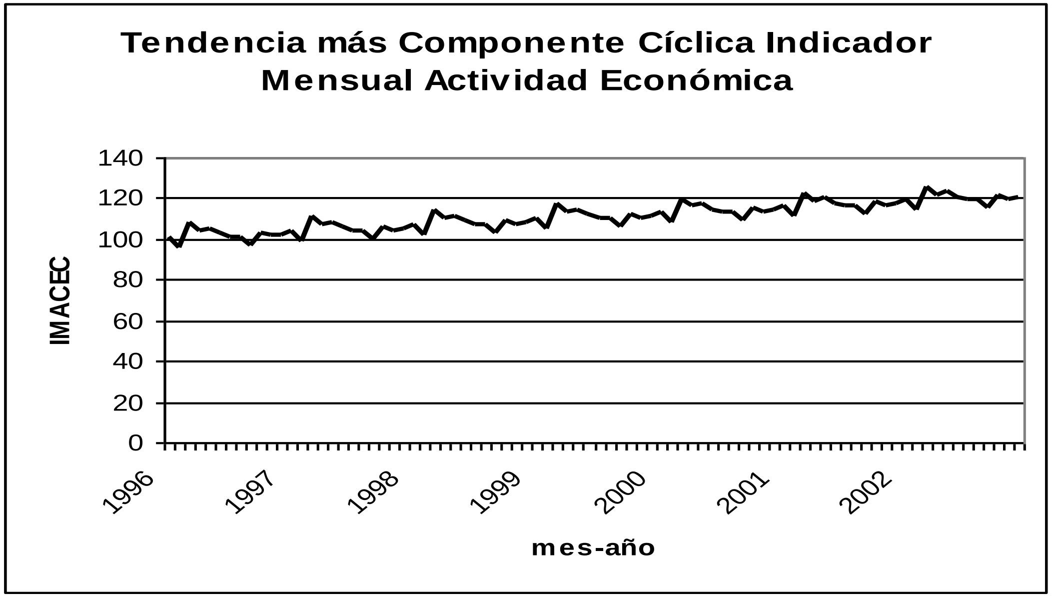
- Entonces para obtener una estimación de la estacionalidad, restamos los valores ajustados de la tendencia a los datos, obteniendo una serie sin tendencia.
- Luego promediamos todos los valores de enero, los de febrero, los de marzo, etc., obteniendo doce valores mensuales promedio:

| Mes | Prom. |
|---|---|
| Enero | 0.8 |
| Febrero | -4.9 |
| Marzo | 6.7 |
| Abril | 2.4 |
| Mayo | 3.8 |
| Junio | 0.8 |
| Julio | -1.2 |
| Agosto | -1.3 |
| Septiembre | -5.4 |
| Octubre | 0.2 |
| Noviembre | -1.7 |
| Diciembre | -1.5 |



Se observan valores altos a partir de marzo, y bajos en torno a septiembre.

Si recomponemos la serie con tendencia y componente cíclica, sin la componente aleatoria, tenemos la situación que se ilustra en el gráfico siguiente:




Con esto se pueden hacer predicciones futuras,

# Introducción Predicción Series Temporales

- Los modelos ARIMA responden al acrónimo de procesos **A**uto**R**regresivos, **I**ntegrados, y Medias móviles (**M**oving **A**verage), y fueron planteados inicialmente por George Box y Gwilym Jenkins en 1970 en su obra "Time Series Analysis: Forecasting and Control (Holden Day, San Francisco, USA)".
- La idea subyacente fundamental consiste en admitir que las series temporales son generadas mediante un Proceso Generador de Datos que puede ser identificado y cuantificado y que, por tanto, pueden ser inferidos sus valores a futuro.
- Cuando realizamos una predicción de la evolución de una determinada serie temporal mediante la descomposición en los componentes estacional, tendencial, cíclico e irregular, el procedimiento consiste en identificar comportamientos regulares a lo largo de la serie (movimientos estacionales, tendenciales y cíclicos ) y extrapolarlos a futuro, asumiendo que los comportamientos irregulares tendrán un efecto promedio nulo.

# INTRODUCCIÓN A LOS MODELOS ARIMA

En el caso de los modelos ARIMA identificaremos igualmente una serie de comportamientos regulares asociados a procesos de evolución temporal conocidos (Procesos de integración, autorregresivos y de Medias móviles) que interactúan con procesos completamente aleatorios (Ruido blanco).

- El proceso de predicción con modelo ARIMA puede, por tanto, resumirse en las siguientes etapas:

1) Identificación de los procesos subyacentes (P.G.D.):

    1.1. Orden de integración

    1.2.- Tipología de procesos AR y MA

2º) Estimación de los coeficientes asociados a los procesos AR y MA.

3º) Validación del modelo estimado.

4º) Cuantificación a futuro de los valores de la serie objetivo.

## INTRODUCCIÓN A LOS MODELOS ARIMA

**Fundamentos :**

La modelización ARIMA asume que toda serie temporal está generada por un proceso estocástico (Proceso Generador de datos PGD) en la que los distintos valores observados $Y_t$ responden a realizaciones (muestras) concretas de un conjunto de N variables aleatorias $Z_t$, que tienen unas determinadas probabilidades de ocurrencia asociadas a sus respectivas funciones de densidad $f(Z_t)$.

Estas funciones de densidad serán, en general, desconocidas y no pueden ser estimadas ya que sólo disponemos de una observación de cada una de ellas, por lo que se hace necesario asumir una serie de simplificaciones para poder realizar cualquier tipo de inferencia estadística.

## Procesos estocásticos elementales: Camino aleatorio.

El *camino aleatorio* es un proceso tal que la diferencia entre dos valores consecutivos de la variable se comporta como un ruido blanco.

$$Z_t - Z_{t-1} = a_t \quad o\, bien \quad Z_t = Z_{t-1} + a_t$$

Si existe una tendencia sistemática en el cambio se denomina *camino aleatorio con deriva*.

$$Z_t - Z_{t-1} = m + a_t \quad o\, bien \quad Z_t = m + Z_{t-1} + a_t$$

## Procesos estocásticos elementales: Proceso Autorregresivo.

Definimos un *proceso autorregresivo de primer orden* AR(1) como un proceso aleatorio que responde a una expresión del tipo

$$Z_t = \rho_0 + \rho_1 Z_{t-1} + a_t \quad o\ bien \quad \breve{Z}_t = \rho_1 \breve{Z}_{t-1} + a_t \quad con\ \breve{Z}_t = Z_t - \rho_0$$

Para que el proceso AR(1) sea estacionario se debe cumplir que $-1<\rho_1<1$, para que $\sigma_z^2$ sea finita y no negativa.

Los procesos autoregresivos pueden generalizarse al orden p AR(p) sin más que añadir términos retardados en la expresión general.

$$Z_t = \rho_0 + \rho_1 Z_{t-1} + \rho_2 Z_{t-2} + \ldots + \rho_p Z_{t-p} + a_t$$

## Procesos estocásticos elementales: Medias móviles.

Definimos una *media móvil de primer orden* MA(1) como un proceso aleatorio que responde a una expresión del tipo

$$Z_t = a_t + \theta_1 a_{t-1} \quad con\ Z_t\ en\ diferencias\ a\ la\ media$$

Los procesos de medias móviles son estacionarios y, al igual que los autoregresivos pueden generalizarse al orden q MA(q) sin más que añadir términos retardados en la expresión general.

$$Z_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \ldots + \theta_q a_{t-q}$$

## Procesos estocásticos elementales: Procesos integrados.

Un proceso *integrado* es aquel que puede convertirse en estacionario aplicando diferencias.

Así, por ejemplo, un camino aleatorio sería un proceso integrado de orden 1 I(1), ya que puede convertirse en estacionario tomando primeras diferencias.

Definimos el orden de integración de un proceso como el número de diferencias que debemos aplicarle para convertirlo en estacionario.

En algunas ocasiones las diferencias deben aplicarse sobre el valor estacional.

$$Z_t - Z_{t-s} = e_t \quad con\ s = 4\ ó\ 12 \quad e_t \quad estacionario$$

## Proceso Generador de Datos.

Mediante la adecuada combinación de estos procesos elementales: integración, AR(p), y MA(q) podemos representar la evolución de cualquier serie temporal.

$$\nabla Y_t = \rho_1 \nabla Y_{t-1} + \rho_2 \nabla Y_{t-2} + \cdots + \rho_p \nabla Y_{t-p} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \cdots + \theta_p a_{t-p}$$

$$\nabla Y_t \rho_p(B) = \theta_q(B) a_t \Rightarrow \nabla Y_t = \frac{\theta_q(B)}{\rho_p(B)} a_t \qquad con \quad \nabla Y_t = Y_t - Y_{t-1} = Y_t(1-B)$$

# Resumen

- Para realizar una predicción es recomendable buscar y analizar la información disponible.
- Para realizar una predicción es necesario argumentarla, independientemente de que dichos argumentos sean correctos o erróneos.
- Es necesario poner a prueba la validez de la predicción.
- Hay que idear un procedimiento experimental, o teórico, para validar la predicción.
- Hay que estar dispuesto a modificar la predicción cuando la evidencia experimental arroja un resultado diferente al propuesto.