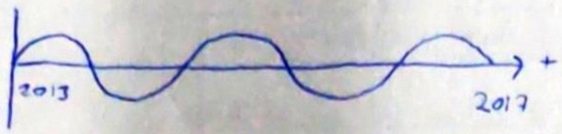
Tema 3: Series temporales

Una serie temporal es un conjunto de datos ordenados cronológicamente. Por ejemplo: Temperaturas máximas registradas en Málaga durante los años 2013-2019.

$\{(1 \text{marzo } 2013, 20°C), (2 \text{marzo } 2013, 20'5°C), (31 \text{ diciembre } 2018, 18°C)\} = \{(t_k, t_k^{max})\}_{k=1}$

Puede verse como una variable bidimensional (t, t^{max})

La devolveremos como $\{x_k\}$

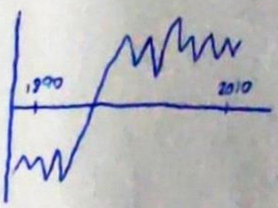


¿Qué observamos en esta serie?

→ Hay "ciclos": las temperaturas suben y bajan según sea verano o invierno

→ A esto se le llama componente estacional de la serie (E_k) (Patrón recursivo o de cambio recurrente)

→ En promedio, no parece que haya habido un incremento de la temperatura, por eso se dice que esta serie es estacionaria.



→ Vemos anomalías de las variaciones de temperatura a lo largo de casi 120 años y observamos un crecimiento de la temperatura, a esta dirección dominante de la serie se llama tendencia (T_k)

→ Puede haber otros valores:

- Componente aleatoria: (variaciones aleatorias, fluctuaciones pequeñas)
- Componentes cíclicas.

Existen muchos modelos matemáticos para analizar una serie temporal.

La idea será:

→ Descomponer en:

Tendencia (T_k), Estacionalidad (E_k), Más ciclos (C_k), Aleatoriedad (A_k)

→ Estudiaremos 3 tipos de modelos:

- Modelo aditivo:

$X_k = T_k + E_k + C_k + A_k$

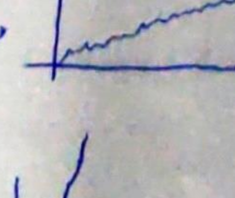
- Modelo multiplicativo:

$X_k = T_k \cdot E_k \cdot C_k \cdot A_k$ →

(Modelo aditivo)

- Modelo mixto:

$X_k = T_k \cdot E_k \cdot C_k + A_k$ →



(Modelo multiplicativo)

Si conozco X_K y estimo T_K $\Rightarrow$ $E_K = X_K - T_K - A_K$ Caso modelo aditivo

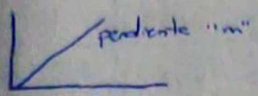
$E_K \approx X_K - T_K$

¿Cómo puedo descomponer la serie?

1) Estimar la tendencia.

¿Cómo? Regresión Lineal. Método gráfico. Filtros

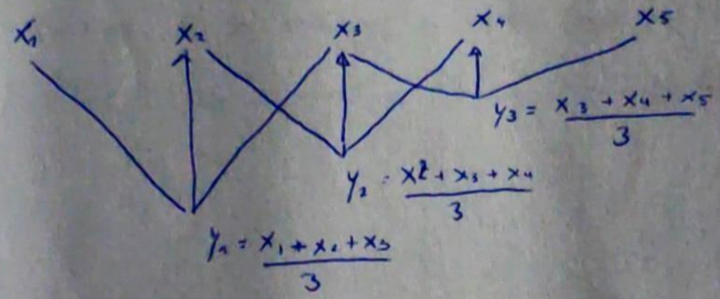
$\frac{(350-320)}{(1970-1980)} = m$



En este curso usaremos la idea de filtro para el cálculo de la tendencia.

Idea filtro: consiste en suavizar la serie temporal, para poder extraer una componente tan fuerte como lo es la tendencia. Un filtro sencillo y efectivo es el de las medias móviles.

Idea de media móvil: $\{X_K\}_{K=1}^{N=5}$ Media móvil orden 3.



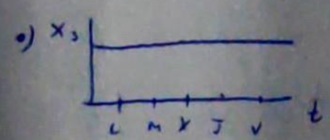
Observación:

- Pierdo información:

Paso de la serie temporal $\{X_K\}$ con 5 datos a la serie temporal de medias móviles $\{Y_K\}_{K=1}^{3}$

En general, si la serie tiene N observaciones, la serie de medias móviles de orden 3 tiene N-2 medias móviles.

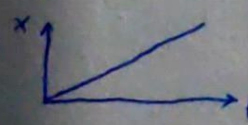
¿Por qué es un filtro?

•) 

Si $X_K = cte$

$Y_2 = \frac{X_1 + X_2 + X_3}{3} = \frac{cte + cte + cte}{3} = cte \Rightarrow Y_K = cte$

•) $X_K = m \cdot K$

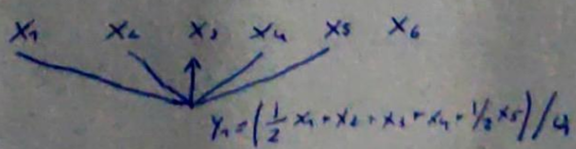


$Y_t = \frac{X_{K-1} + X_K + X_{K+1}}{3} = m(K-1) + mt + m(K+1) = mK \Rightarrow Y_K = mK$

En general, una media móvil de orden "K impar" es:

$Y_i = \frac{X_{i-s} + X_{i-s+1} + \dots + X_i + \dots + X_{i+s}}{K}$ } K términos de X_i

Calcular la media móvil de orden $K=4$ de los puntos:



En lugar de calcular la media de orden 4, voy a calcular la media móvil de orden 5, pero con distintos pasos.

Ejemplo: Un productor estudio el precio de venta a lo largo de 4 años obteniendo la siguiente serie

	2012			2013			2014			2015		
	I	II	III	I	II	III	I	II	III	I	II	III
Precio $= X_k$	5'5	6'3	6'2	6	7'6	7'4	5'8	6'1	7'2	6'2	7'8	7'7
T_k	—	6	6'1167	6'6	7	6'9333	6'9333	6'3667	6'5	7'0667	7'233	—
$E_k \cdot A_k$	—	1'05	1'0054	0'9090	1'0857	1'0673	0'9016	0'9581	1'1077	0'8774	1'0783	—
$E_k \cdot A_k / J_i'$	—	1'0065	—									—

$I_1 = 0'8960$ $\qquad$ $J_2 = 1'0430$ $\qquad$ $I_3 = 1'0604$

debe ocurrir que $\frac{I_1 + I_2 + I_3}{3} = 1$

Inciso:

$X_k = T_k E_k A_k$, suele admitirse que $\{T_k\}$ y $\{E_k A_k\}$ son independientes

$\frac{I_1 + I_2 + I_3}{3} = 0'9998$

Corrección de índices: $I_1' = 1/0'9998 \cdot I_1 = 0'8962$

$I_2' = 1/0'9998 \cdot I_2 = 1'0432$ $\qquad \frac{I_1' + I_2' + I_3'}{3} = 1$

$I_3' = 1/0'9998 \cdot I_3 = 1'0606$

Interpretación de índices de estacionalidad.

Si me fijo en el primer índice $I_1' = 0'8962 \Rightarrow$ se interpreta como que en ese trimestre, el primero, el valor del precio es un 10'38% inferior a la media $(100 - 100 \cdot 0'8962)$

Si dividimos:

$\frac{X_k}{T_k} = E_k A_k$ entre los índices I_i' entonces: $\frac{X_k}{T_k I_i'} = \frac{E_k A_k}{I_i'} = A_k \cdot 1$

$E_k = \frac{E_k \cdot A_k}{A_k}$

$A_2 = 1'0065$. Interpretación: En el segundo trimestre de 2012, los factores aleatorios hicieron que el precio aumentase un 0'65% $(100 \cdot 1'0065 - 100)$

$X_k = (T_k E_k) A_k$ $\qquad$ $X_2 = T_2 E_2 \cdot 1'0065$

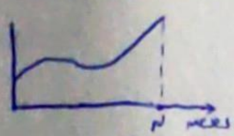
En resumen:

- Saco la tendencia T_k, usando medias móviles, de orden $k=3$ (porque trabajo con trimestres)
- Saco $E_k A_k$, dividiendo X_k/T_k (Modelo multiplicativo $X_k = T_k E_k A_k$)
- Averiguar la componente aleatoria A_k
  - → Para ello recurro a los índices de estacionalidad (medias de $E_k A_k$ por trimestres (ciclos))
  - → Problema: al calcularlos a mano, cometo errores y debo corregirlos para que la media de los índices sea 1.
  - → Una vez tenga corregidos I_k', divido: $\frac{E_k \cdot A_k}{I_k'}$
- y con eso obtengo A_k
- $X_k = T_k E_k A_k \longrightarrow E_k = X_k / T_k A_k$

Todo lo que hemos visto hasta ahora de series temporales nos sirven para entender la serie (para explicar un fenómeno sometido a estudio), o también nos podría servir para relacionar 2 series (o dos fenóm)

Otra utilidad importante de las series temporales es la de tratar de predecir futuros resultados de esa serie.

Por ejemplo si tengo $\{X_k\}_{k=1}^{N}$



Me puede interesar conocer el próximo valor de la serie X_{N+1}

¿Cómo podría hacerlo?

Si tengo un modelo multiplicativo $X_k = T_k E_k A_k \Rightarrow X_{N+1} \approx T_{N+1} \cdot E_{N+1} \cdot 1$

En ese caso necesito saber T_{N+1} y E_{N+1}, pero T_{N+1} y E_{N+1} los voy a inferir a partir de $\{x_1, x_2 \dots x_N\}$

- T_{N+1}

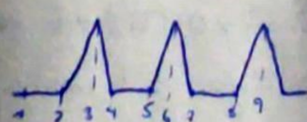
  Calculo la recta de regresión lineal $\hat{T}(t)$ usando los puntos $\{(t_1, T_1), (t_2, T_2), (t_3, T_3) \dots (t_N, T_N)\}$

  y entonces $T_{N+1} \approx \hat{T}(t_{N+1})$

- E_{N+1} (Ejemplo de ciclos de 3)

  $\underbrace{E_1, E_2, E_3}_{\text{Ciclo de 3}}, \underbrace{\overset{E_1}{E_4}, \overset{E_2}{E_5}, \overset{E_3}{E_6}}_{\text{Ciclo de 3}}$

El trabajo recae en averiguar la longitud de los ciclos (en este ejemplo tenemos ciclos de 3)

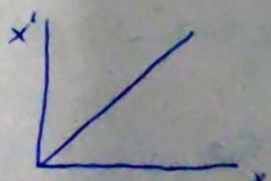


Considero: $\{(X_k, X_{k+2})\}_{k=1}^{N-2}$

Si yo calculo el coeficiente de correlación lineal $r = \frac{\mathrm{Cov}(X, X^L)}{\sigma_X \, \sigma_{X^L}}$ serie desplazada L lugares

Entonces:

Si $r \approx 1 \Rightarrow X$ y X^L están correlados.



$\Longrightarrow$