# Series de Tiempo y Pronósticos

César Andrés Ojeda Echverri
cesar.ojeda@correounivalle.edu.co

Universidad del Valle
Escuela de Estadística

Febrero, 2021

## Texto Guía

Wei, W. W. S. (2006), *Time Series Analysis: Univariate and Multivariate Methods* (2nd ed.), Pearson Addison Wesley.

## Textos complementarios

- Box, G. E. P., Jenkins, G. M., Reinsel, G. C. and Ljung, G.M. (2016), *Time Series Analysis: Forecasting and Control* (5th ed.), Wiley Series in Probability and Statistics, Wiley.
- Brockwell, P. J., and Davis, R. A. (2016), *Introduction to Time Series and Forecasting* (3rd ed.), Springer Text in Statistics, Springer-Verlag.
- Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods* (2nd ed.), Springer Text in Statistics, Springer.

## Textos complementarios


- Cryer, J. D. and Chan, K. (2008), *Time Series Analysis with Applications in R* (2nd ed.), Springer.
- Hamilton, J. D. (1994), *Time Series Analysis*, Princeton University Press.
- Priestley, M. B. (1981), *Spectral Analysis and Time Series*, Academic Press.
- Fuller, W. A. (1996), *Introduction to Statistical Time Series* (2nd ed.), Wiley Series in Probability and Statistics, Wiley.
- Shumway, R. H. and Stoffer, D. S. (2017), *Time Series Analysis and Its Applications with R Examples* (4th ed.), Springer Text in Statistics, Springer.
- Cowpertwait, P. S. P. and Metcalfe, A. V. (2009), *Introductory Time Series with R*, Springer.

## Textos complementarios

- Peña, D. (2005), *Análisis de Series Temporales*, Alianza Editorial.
- Guerrero, V. M. (2003), *Análisis Estadístico de Series de Tiempo Económicas* (2da ed.), Thomson.
- Diaz, R. N. (1995), *Modelación de Procesos Estocásticos y Pronósticos de Series Temporales*, Publicaciones Facultad de Ingeniería.

## Programas Estadísticos

- R
  http://www.r-project.org/
- Time Series Modelling (TSM)
  http://www.timeseriesmodelling.com/
- Scientific Computing Associates (SCA)
  http://www.scausa.com/
- ITSM2000
  http://www.math.ku.dk/~sjo/itsm2000/

## Evaluación

- Taller 1: 35 %.
- Taller 2: 35 %.
- Presentación final: 30 %.

**Importante:** Los talleres eventualmente se deben sustentar oralmente por parte de los integrantes.

# Introducción

- Informalmente, una serie de tiempo es una sucesión ordenada de observaciones. Generalmente, este ordenamiento se hace a través del tiempo y en intervalos igualmente espaciados. Sin embargo, existen situaciones donde el ordenamiento se presenta en otras dimensiones tales como el espacio.
- Las series de tiempo tienen aplicación en muchas áreas. Por ejemplo:
    - **En agricultura:** La producción anual de las cosechas y sus precios de mercado.
    - **En economía y negocios:** Los precios de cierre diarios de las acciones, las tasas de interés semanales, los índices de precios mensuales, las ventas trimestrales, las ganancias anuales, las tasas de desempleo.

# Introducción

- También:
  - **En ingeniería:** Las ondas de sonido, señales eléctricas, el voltaje.
  - **En geofísica:** Registros de turbulencia tales como las olas del mar y el ruido de la tierra en una zona específica.
  - **En medicina:** Trazados de electroencefalogramas (EEG) y electrocardiogramas (EKG), tiempos de supervivencia a una enfermedad.
  - **En meteorología:** La velocidad del viento horaria, la temperatura diaria, la lluvia anual.
  - **En hidrología:** Los caudales de los ríos.
  - **En control de calidad:** La evolución de un proceso de producción de acuerdo a un cierto valor objetivo.
  - **En las ciencias sociales:** Las tasas de nacimiento anuales, las tasas de mortalidad anuales, tasas de accidentes semanales, tasas de criminalidad.

# Introducción

- Las series de tiempo pueden ser:
  - **Discretas:** Las cuales son observadas en intervalos específicos de tiempo tales como horas, días, semanas, meses, etc. Por ejemplo, tasas de interés, rendimientos o volumen de ventas.
  - **Continuas:** Las cuales son observadas de forma continua en el tiempo. Por ejemplo, el electrocardiograma.

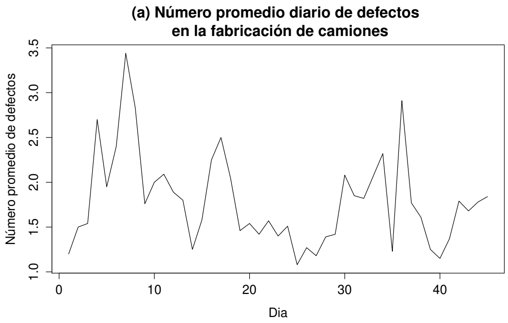
¡El propósito de este curso es el estudio de series de tiempo discretas!

# Introducción

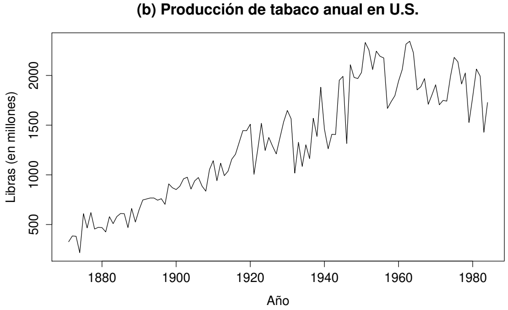
- El estudio de una serie de tiempo tiene varios objetivos:
  - Comprender y describir el mecanismo que genera las observaciones.
  - Pronosticar valores futuros.
  - Realizar control óptimo de un sistema.

# Introducción

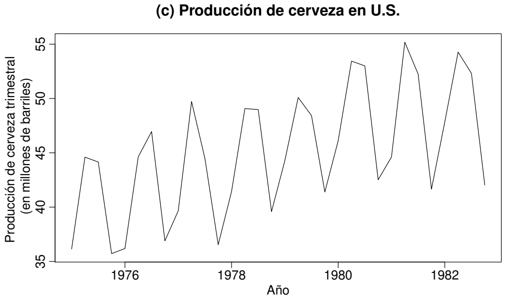
- La naturaleza intrínseca de las series de tiempo implica que sus observaciones son dependientes o correlacionadas, y por tanto el orden de las mediciones es importante. Las técnicas estadísticas que se basan en muestras aleatorias (independencia de las observaciones) no son aplicables en el análisis y se necesita construir nuevos métodos.



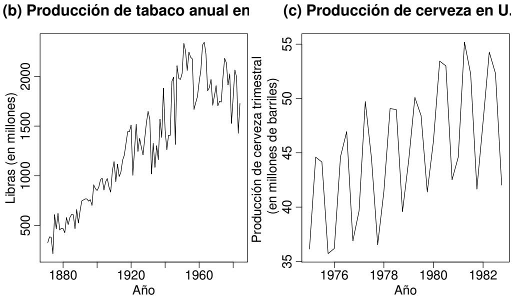

- El número promedio diario de defectos por camión encontrados al final de la linea de ensamble de fabrica camiones parece variar alrededor de un nivel fijo.
- Las series de tiempo que exhiben este fenómeno se dice que son estacionarias en media y son casos especiales de series de tiempo estacionarias.



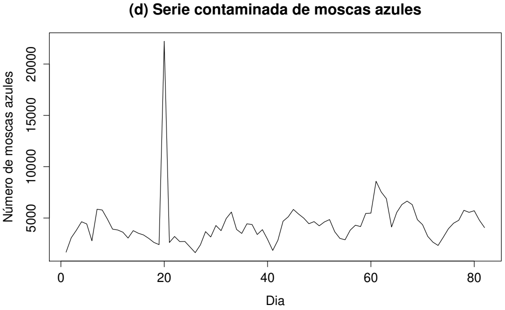

- Por otro lado, la producción de tabaco anual en E.U. no parece variar alrededor de un nivel fijo y exhibe una tendencia a crecer en términos generales.
- Además, la varianza de esta serie incrementa a medida que incrementa el nivel.
- Las series de tiempo que exhiben este fenómeno se dicen ser no estacionarias en media, ni en varianza y constituye un ejemplo de series de tiempo no estacionarias.




- La producción trimestral de cerveza en U.S. presenta otro patrón característico cuya naturaleza es repetitiva debido a la variación estacional.



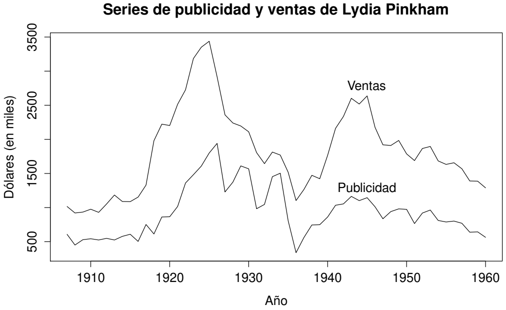

- Series de tiempo no estacionarias tales como (b) y (c) pueden reducirse a series estacionarias mediante transformaciones adecuadas.




- La serie de contaminada de moscas azules presenta otro fenómeno de no estacionariedad debido a un cambio en la estructura de la serie a partir de alguna perturbación externa.
- Este tipo de no estacionariedad no puede ser removido por transformaciones estándar.
- Tales perturbaciones externas son denominadas intervenciones o outliers.

# Introducción

- Las series de tiempo son a menudo afectadas por dichas intervenciones o outliers.
- Por ejemplo:
  - Las series de producción son afectadas por huelgas.
  - Las series mal registradas debido a fallas en las maquinas.
- Para modelar tales perturbaciones externas existen los modelos de intervención y de outliers.




- Los datos en series de tiempo, a menudo involucran observaciones simultaneas de varias variables.
- En los negocios, generalmente se intenta promover las ventas mediante la publicidad.
- Por tanto, parece natural construir un modelo dinámico que relacione las ventas actuales con los valores presentes y pasados de la publicidad para estudiar su efecto en las ventas y mejorar sus pronósticos.

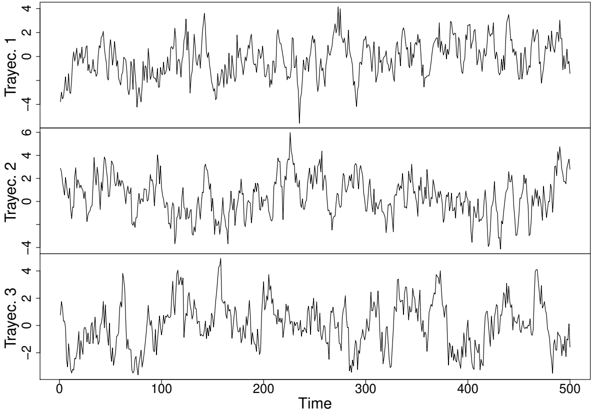
# Conceptos Fundamentales

- Los procesos estocásticos son el marco teórico formal para el estudio de las series de tiempo.
- **Proceso estocástico:** Un proceso estocástico es una familia de variables aleatorias indexadas en el tiempo $Z(w, t)$, donde $w$ pertenece a un espacio muestral y $t$ pertenece a un conjunto de índices.
  - Para un valor fijo de $t$, $Z(w, t)$ es un variable aleatoria.
  - Para un $w$ determinado, $Z(w, t)$, como función de $t$, es llamada función muestral o realización.

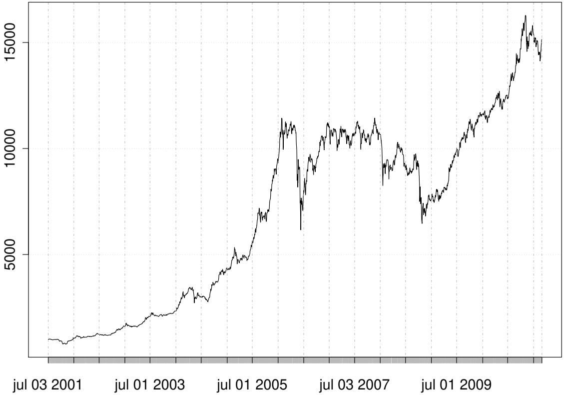
Asumiremos que el conjunto de índices
es el conjunto de los enteros.

# Conceptos Fundamentales

- La población es el conjunto de todas las posibles realizaciones de un proceso estocástico.
- Formalmente, una serie de tiempo es una realización o función muestral de cierto proceso estocástico y corresponde a una sola observación del proceso.




- Ejemplo de tres trayectorias de un proceso estocástico dado.




- Realización diaria del Índice General de la Bolsa de Valores de Colombia entre 2000-07-03 y 2011-03-03.

# Conceptos Fundamentales

- **Definición:** Un modelo de series de tiempo para un conjunto de observaciones $\{x_t\}$ es una especificación de las distribuciones conjuntas (o posiblemente sólo de las medias y covarianzas) de una sucesión de variables aleatorias $\{X_t\}$, de las cuales $\{x_t\}$ se postula a ser una realización.

# Conceptos Fundamentales

- **Caracterización de los procesos estocásticos:** Considere un conjunto finito de variables aleatorias $\{Z_{t_1}, Z_{t_2}, \ldots, Z_{t_n}\}$ de un proceso estocástico $\{Z(w,t) : t = 0, \pm 1, \pm 2, \ldots\}$. La función de distribución $n$-dimensional se define como:

$$F_{Z_{t_1},\ldots,Z_{t_n}}(x_1,\ldots,x_n) = P\{w : Z_{t_1} \leq x_1, \ldots, Z_{t_n} \leq x_n\},$$

donde $x_i$, para $i = 1, \ldots, n$, son constantes reales.

# Conceptos Fundamentales

- Un proceso se dice ser estacionario de primer orden en distribución si su función de distribución unidimensional es invariante en el tiempo, es decir:

$$F_{Z_{t_1}}(x_1) = F_{Z_{t_1+k}}(x_1),$$

para cualquier entero $t_1$, $k$, y $t_1 + k$.

- Un proceso se dice ser estacionario de segundo orden en distribución si:

$$F_{Z_{t_1},Z_{t_2}}(x_1, x_2) = F_{Z_{t_1+k},Z_{t_2+k}}(x_1, x_2),$$

para cualquier entero $t_1$, $t_2$, $k$, $t_1 + k$, y $t_2 + k$.

# Conceptos Fundamentales

- Un proceso se dice ser estacionario en distribución de orden $n$ si:

$$F_{Z_{t_1},\ldots,Z_{t_n}}(x_1,\ldots,x_n) = F_{Z_{t_1+k},\ldots,Z_{t_n+k}}(x_1,\ldots,x_n) \qquad (1)$$

para cualquier $n$-dupla $(t_1,\ldots,t_n)$ y $k$ enteros.

Un proceso se dice ser estrictamente estacionario (o fuertemente estacionario, o completamente estacionario) si (1) se cumple para cualquier $n$, es decir, $n = 1, 2, \ldots$

**Observación:** Si (1) es cierto para $n = m$, entonces también es cierto para $n \leq m$, es decir, la estacionariedad de orden alto implica siempre la estacionariedad de ordenes más bajos.

# Conceptos Fundamentales

**Ejemplo:** Un ejemplo trivial de un proceso estocástico estrictamente estacionario lo constituye una sucesión de variables aleatorias independientes e idénticamente distribuidas (*i.i.d.*). Sin embargo, este tipo de series no son de interés en las series de tiempo.

# Conceptos Fundamentales

- **Notación:** Entendiendo que un proceso estocástico, $Z(w, t)$, es un conjunto de variables aleatorias indexadas en el tiempo definidas sobre un espacio muestral, generalmente se omitirá la variable $w$ y simplemente $Z(w, t)$ se escribirá como $Z_t$ o $Z(t)$.

Un proceso estocástico es llamado proceso de valor real si este asume solamente valores reales. A menos que se indique lo contrario, los proceso discutidos en este curso serán procesos de valor real.

# Conceptos Fundamentales

Para un proceso de valor real dado $\{Z_t : t = 0, \pm 1, \pm 2, \ldots\}$, se define:

- La función media del proceso: $\mu_t = E(Z_t)$,
- La función de varianza del proceso: $\sigma_t^2 = E(Z_t - \mu_t)^2$,
- La función de covarianza entre $Z_{t_1}$ y $Z_{t_2}$:

$$\gamma(t_1, t_2) = E(Z_{t_1} - \mu_{t_1})(Z_{t_2} - \mu_{t_2}),$$

- La función de correlación entre $Z_{t_1}$ y $Z_{t_2}$:

$$\rho(t_1, t_2) = \frac{\gamma(t_1, t_2)}{\sqrt{\sigma_{t_1}^2}\sqrt{\sigma_{t_2}^2}}.$$

Para un proceso estrictamente estacionario, dado que la función de distribución es la misma para todo $t$, se cumple que:

- La función de media del proceso: $\mu_t = \mu$ para todo $t$, siempre que $E(|Z_t|) < \infty$.
- La función de varianza del proceso: $\sigma_t^2 = \sigma^2$ para todo $t$, siempre que $E(Z_t^2) < \infty$.
- La función de covarianza entre $Z_{t_1}$ y $Z_{t_2}$:

  $$\gamma(t_1, t_2) = \gamma(t_1 + k, t_2 + k),$$

  y la función de correlación sería:

  $$\rho(t_1, t_2) = \rho(t_1 + k, t_2 + k),$$

  dado que $F_{Z_{t_1}, Z_{t_2}}(x_1, x_2) = F_{Z_{t_1+k}, Z_{t_2+k}}(x_1, x_2)$ para cualquier entero $t_1$, $t_2$ y $k$.

Para un proceso estrictamente estacionario que posee los primeros dos momentos finitos, la covarianza y la correlación entre $Z_t$ y $Z_{t+k}$ depende sólo de la diferencia entre los tiempos, es decir, $k$.

- Sea $t_1 = t - k$ y $t_2 = t$, por tanto:

  $$\gamma(t_1, t_2) = \gamma(t - k, t) = \gamma(t, t + k) = \gamma_k,$$

  es decir, la función solamente depende de la diferencia entre los tiempos que separan a $Z_{t_1}$ y $Z_{t_2}$.
- De manera similar para la función de correlación se cumple que:

  $$\rho(t_1, t_2) = \rho(t - k, t) = \rho(t, t + k) = \rho_k,$$

  la cual también dependen sólo de la diferencia entre los tiempos.

En la práctica, generalmente es muy difícil probar si un proceso es estrictamente estacionario, y se trata de caracterizar los procesos estocásticos en términos de sus momentos cuyas propiedades se pueden verificar de manera más sencilla.

- Un proceso se dice ser débilmente estacionario de orden $n$ si todos sus momentos conjuntos hasta de orden $n$ existen y son invariantes en el tiempo.
- Un proceso débilmente estacionario de segundo orden tendrá medias y varianzas constantes, con covarianzas y correlaciones como funciones solamente de la diferencia entre los tiempos. Esta clase de procesos es también llamado proceso estacionario en sentido amplio o proceso estacionario en covarianza o simplemente estacionario.

# Conceptos Fundamentales

**Observación:** Un proceso estrictamente estacionario con los dos primeros momentos finitos es también un proceso débilmente estacionario de segundo orden o estacionario en covarianza.

**Observación:** Sin embargo, un proceso estrictamente estacionario puede no ser estacionario en covarianza, ya que puede no tener momentos de primer y segundo orden finitos. Por ejemplo, el proceso formado por una sucesión de variables aleatorias *i.i.d.* Cauchy.

# Conceptos Fundamentales

**Ejemplo:** Considere la sucesión en el tiempo:

$$Z_t = A \sin(wt + \theta)$$

donde $A$ es una variable aleatoria de media cero y varianza uno, y $\theta$ es otra variable aleatoria con distribución uniforme continua entre $-\pi$ y $\pi$ independiente de $A$. Pruebe que el proceso estocástico es débilmente estacionario.

**Recuerde que:**

$$\sin(\lambda)\sin(u) = \frac{1}{2}\left[\cos(\lambda - u) - \cos(\lambda + u)\right].$$

# Conceptos Fundamentales

**Ejemplo:** Sea $Z_t$ una sucesión de variables aleatorias independientes alternadas inicialmente con una distribución normal estándar $N(0, 1)$ y a continuación una distribución uniforme discreta que toma sólo dos valores ($1$ o $-1$) con igual probabilidad ($1/2$). Muestre que el proceso es estacionario en covarianza, sin embargo no es estrictamente estacionario.

# Conceptos Fundamentales

- En la práctica, generalmente se trabaja con procesos estocásticos estacionarios en covarianza. Este es un supuesto menos restrictivo que la estacionariedad estricta y más fácil de probar.

# Conceptos Fundamentales

- Un proceso estocástico se dice ser un proceso normal o gaussiano si sus distribuciones de probabilidad conjuntas son normales.
- Dado que una distribución normal es únicamente caracterizada por sus primeros dos momentos, entonces la estacionariedad estricta y la estacionariedad débil son equivalentes en los procesos gaussianos.

**Nota:** A menos que se indique lo contrario, se asume que los proceso estocásticos que se estudian son gaussianos.

# Conceptos Fundamentales

- Como en otras áreas de la estadística, la mayor parte de los resultados en series de tiempo son establecidos para procesos gaussianos.
- Así, las funciones de autocorrelación y las funciones de autocorrelación parcial definidas a continuación se convierten en herramientas fundamentales en el análisis de series de tiempo.

Para un proceso estacionario $\{Z_t\}$, se tiene que $E(Z_t) = \mu$ y que $Var(Z_t) = E(Z_t - \mu)^2 = \sigma^2$, las cuales son constantes, y además que $Cov(Z_t, Z_s)$ son sólo funciones de la diferencia entre los tiempo $|t - s|$.

- Por tanto, en este caso, la covarianza entre $Z_t$ y $Z_{t+k}$ es:

$$\gamma_k = Cov(Z_t, Z_{t+k}) = E(Z_t - \mu)(Z_{t+k} - \mu).$$

- La correlación entre $Z_t$ y $Z_{t+k}$ es:

$$\rho_k = \frac{Cov(Z_t, Z_{t+k})}{\sqrt{Var(Z_t)}\sqrt{Var(Z_{t+k})}} = \frac{\gamma_k}{\gamma_0},$$

donde $Var(Z_t) = Var(Z_{t+k}) = \gamma_0$.

**Nota:** Como función de $k$, $\gamma_k$ es llamada función de autocovarianza y $\rho_k$ es llamada función de autocorrelación (ACF).

# Conceptos Fundamentales

**Propiedades:**

1. $\gamma_0 = Var(Z_t)$; $\rho_0 = 1$.
2. $|\gamma_k| \leq \gamma_0$; $|\rho_k| \leq 1$.
3. $\gamma_k = \gamma_{-k}$ y $\rho_k = \rho_{-k}$ para todo $k$, es decir, $\gamma_k$ y $\rho_k$ son funciones pares, y por tanto simétricas alrededor del rezago $k = 0$. Dada esta simetría, no es necesario calcular las funciones para todo entero $k$. Es suficiente calcularlas para $k > 0$.
4. Las funciones de autocovarianzas y autocorrelaciones son semidefinidas positivas, en el sentido de que para cualquier conjunto de puntos en el tiempo $t_1, t_2, \ldots, t_n$, y para cualquier conjunto de números reales $\alpha_1, \alpha_2, \ldots, \alpha_n$,
    4.1 $\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \gamma_{|t_i - t_j|} \geq 0$,
    4.2 $\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \rho_{|t_i - t_j|} \geq 0$.

# Conceptos Fundamentales

En efecto, sea $X = \sum_{i=1}^{n} \alpha_i Z_{t_i}$. Entonces:

$$0 \leq Var(X) = Var(\sum_{i=1}^{n} \alpha_i Z_{t_i}) =$$

$$\sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j Cov(Z_{t_i}, Z_{t_j}) = \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j \gamma_{|t_i - t_j|}.$$

- **Nota:** No toda función arbitraria que satisface las condiciones de 1 a 3 es una función de autocovarianza o autocorrelación para un proceso.
- **Nota:** Una condición necesaria para que una función sea función de autocovarianza o autocorrelación de cualquier proceso es que esta sea semidefinida positiva.

# Conceptos Fundamentales

**Ejemplo:** Considere un proceso estocástico que tiene la siguiente función de autocorrelación:

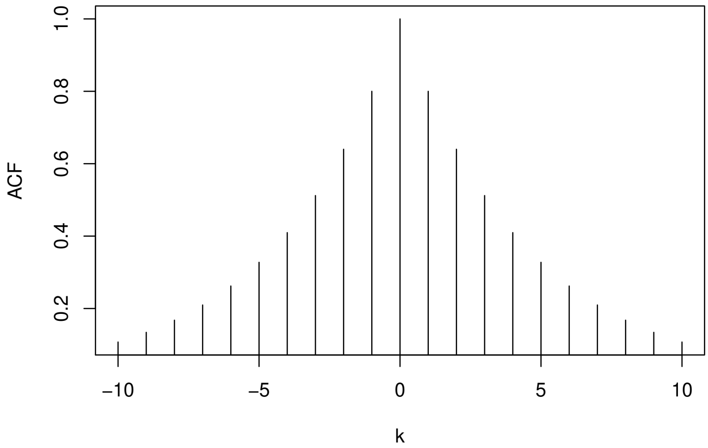
$$\rho_k = \phi^{|k|}, k = 0, \pm 1, \pm 2, \ldots$$

donde $|\phi| < 1$.

# Conceptos Fundamentales

Caso 1: $0 < \phi < 1$. Considere $\phi = 0.8$.

| k | ACF | k | ACF | k | ACF | k | ACF | k | ACF |
|---|---|---|---|---|---|---|---|---|---|
| -10 | 0.11 | -5 | 0.33 | 0 | 1.00 | 5 | 0.33 | 10 | 0.11 |
| -9 | 0.13 | -4 | 0.41 | 1 | 0.80 | 6 | 0.26 | | |
| -8 | 0.17 | -3 | 0.51 | 2 | 0.64 | 7 | 0.21 | | |
| -7 | 0.21 | -2 | 0.64 | 3 | 0.51 | 8 | 0.17 | | |
| -6 | 0.26 | -1 | 0.80 | 4 | 0.41 | 9 | 0.13 | | |

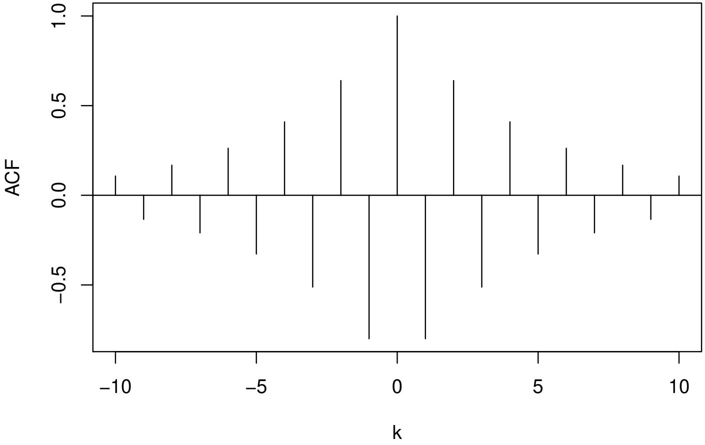



- El gráfico de $\rho_k$ contra $k$ es llamado correlograma.

# Conceptos Fundamentales

Caso 2: $-1 < \phi < 0$. Considere $\phi = -0.8$.

| k | ACF | k | ACF | k | ACF | k | ACF | k | ACF |
|---|---|---|---|---|---|---|---|---|---|
| -10 | 0.11 | -5 | -0.33 | 0 | 1.00 | 5 | -0.33 | 10 | 0.11 |
| -9 | -0.13 | -4 | 0.41 | 1 | -0.80 | 6 | 0.26 | | |
| -8 | 0.17 | -3 | -0.51 | 2 | 0.64 | 7 | -0.21 | | |
| -7 | -0.21 | -2 | 0.64 | 3 | -0.51 | 8 | 0.17 | | |
| -6 | 0.26 | -1 | -0.80 | 4 | 0.41 | 9 | -0.13 | | |




- Dada la simetría de la ACF con respecto a cero, en ambos casos basta realizar la gráfica con $k > 0$.