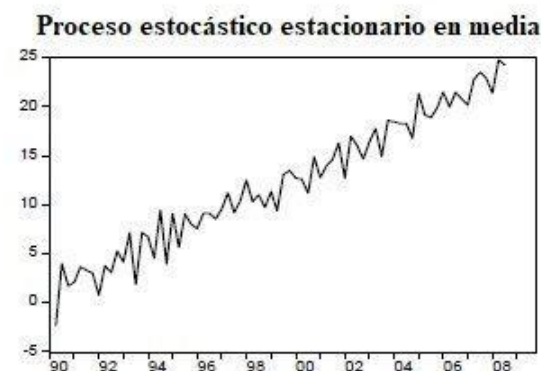
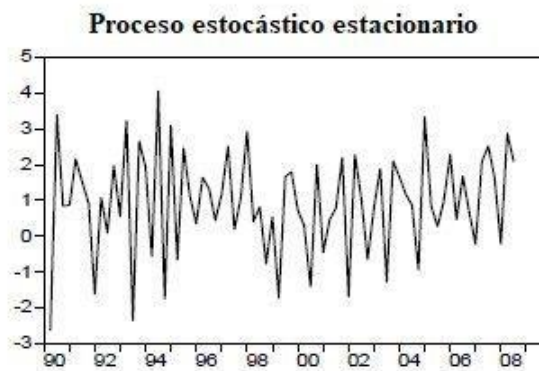
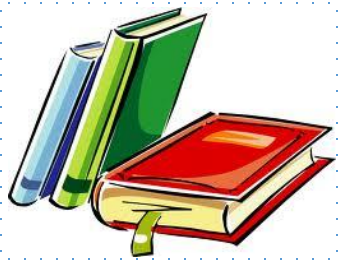


UNIDAD V

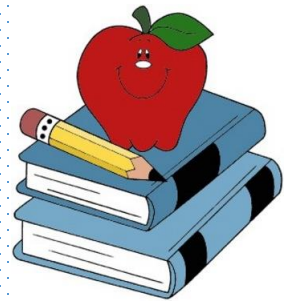
PROCESOS PARTICULARES



DOCENTE: ING. IVETT JACQUELINE TANCARA ZAMBRANA



TEMARIO



1. Modelos lineales en tiempo

2. Proceso AR1

3. Ruido Blanco

4. Procesos Gaussianos

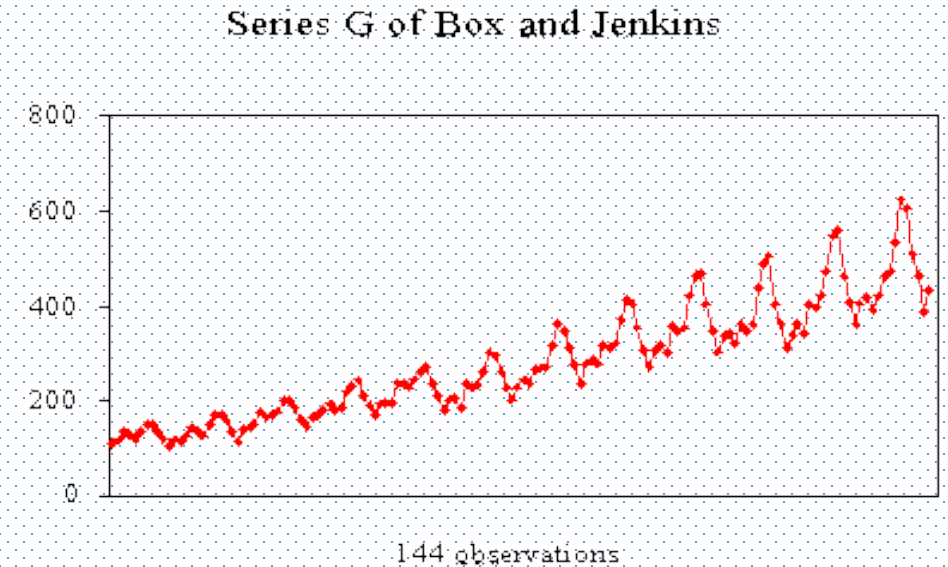
5. Pruebas no paramétricas

6. Ecuación de Chapman - Kolmogorov

1. MODELOS LINEALES EN TIEMPO

Se han analizado las series temporales desde un punto de vista determinístico o clásico. A partir de ahora se estudian desde un punto de vista estocástico o moderno, que utiliza métodos mas complejos y su aplicación requiere series mas largas.

Box y Jenkins han desarrollado modelos estocásticos para series temporales que tienen en cuenta la dependencia existente entre los datos, esto es; cada observación en un momento dado es modelada en función de las anteriores.



1. MODELOS LINEALES EN TIEMPO

Los análisis se basan en un modelo explícito. Los modelos se conocen con el nombre genérico de **ARIMA** (Autoregressive Integrated Moving Average – Media móvil integrada autoregresiva), que deriva de sus tres componentes **AR** (Autoregresivo), **I** (Integrado) y **MA** (Medias móviles).

El **MODELO ARIMA** permite describir un valor como una función lineal de datos anteriores y errores debidos al azar, además, puede incluir un componente cíclico o estacional. Es decir, debe contener todos los elementos necesarios para describir el fenómeno.

Box y Jenkins recomiendan como mínimo **50 observaciones** en la serie temporal.

1. MODELOS LINEALES EN TIEMPO

❖ FASES DEL MODELO ARIMA

La metodología de Box y Jenkins se resume en cuatro fases:

1. Primera fase

Identificar el posible modelo ARIMA que sigue la serie, lo que requiere:

- Decidir que transformaciones aplicar para convertir la serie observada en una serie estacionaria.
- Determinar un modelo ARMA para la serie estacionaria, es decir; los ordenes p y q de su estructura autorregresiva y de media móvil.

1. **MODELOS LINEALES EN TIEMPO**

2. Segunda fase

Seleccionado provisionalmente un modelo para la serie estacionaria, se pasa a la segunda etapa de estimación, donde los parámetros **AR** y **MA** del modelo se estiman por máxima verosimilitud y se obtienen sus errores estándar y los residuos del modelo.

3. Tercera fase

Es el diagnóstico, donde se comprueba que los residuos no tienen estructura de dependencia y siguen un proceso de ruido blanco. Si los residuos muestran estructura se modifica el modelo para incorporarlo y repetir las etapas anteriores hasta obtener un modelo adecuado.

1. MODELOS LINEALES EN TIEMPO

4. Cuarta fase

Es la predicción, una vez que se ha obtenido un modelo adecuado se realizan predicciones con el mismo.

❖ PASOS A SEGUIR PARA EL ANALISIS DE DATOS

- a) **Recolección de datos:** Es conveniente disponer de 50 o mas dato, y en el caso de series mensuales trabajar entre 6 y diez años completos.
- b) **Representación gráfica:** Es de gran utilidad disponer de un gráfico de la serie para decidir sobre la estacionariedad.

1. MODELOS LINEALES EN TIEMPO

- c) **Transformación previa de la serie:** Cuando la serie no es estacionaria en varianza se requiere una transformación logarítmica. No obstante la transformación logarítmica es muy frecuente incluso en series con dispersión relativamente constantemente en el tiempo.
- d) **Eliminación de la tendencia:** La observación del gráfico de la serie indica la existencia o no de tendencia. Una tendencia lineal será corregida tomando primeras diferencias, que será el caso más frecuente.

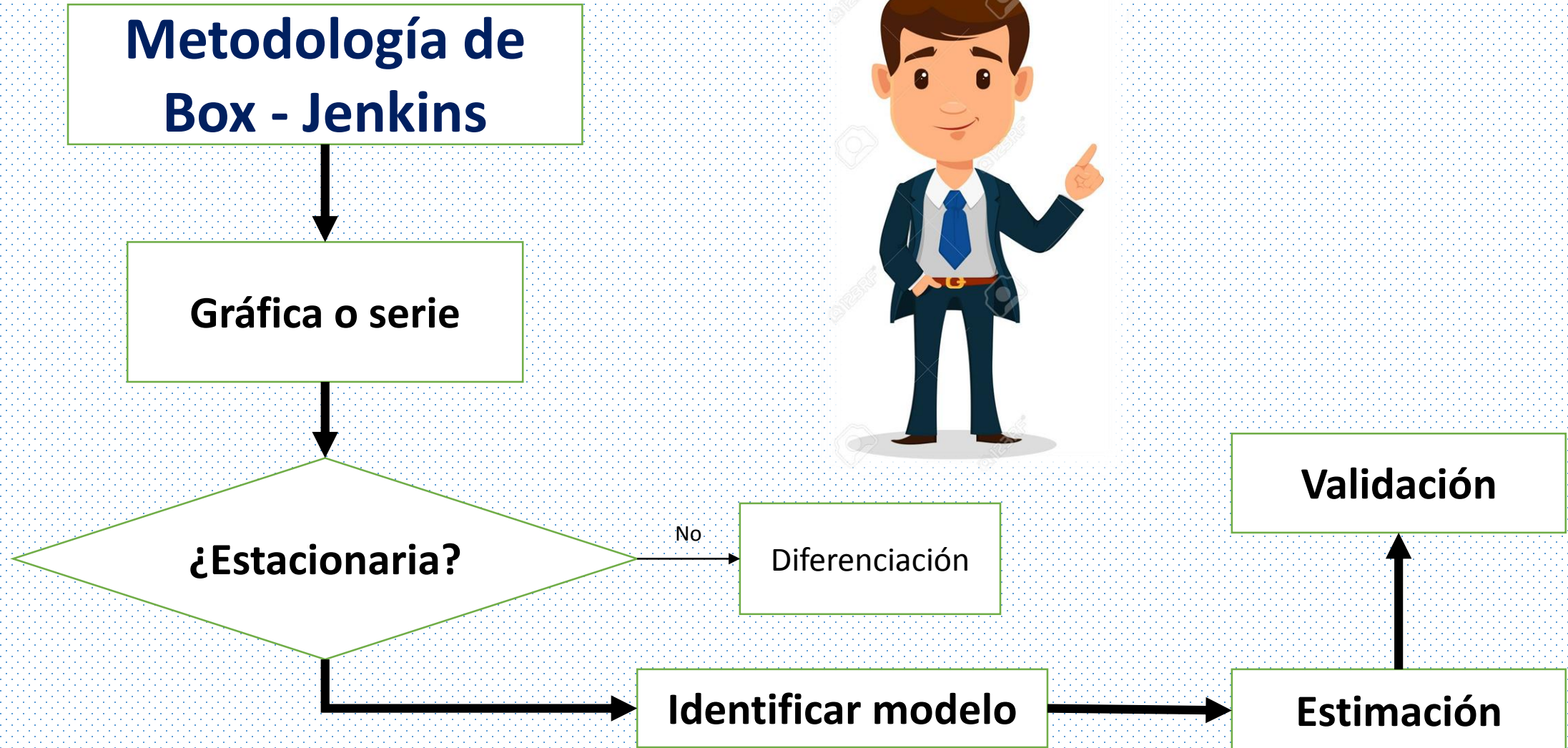
1. MODELOS LINEALES EN TIEMPO

- e) **Identificación del modelo:** Consiste en determinar el tipo de modelo mas adecuado, esto es, el orden de los procesos autoregresivos y de medias móviles de los componentes regular y estacional.
- f) **Estimación de los coeficientes del modelo:** Decidido el modelo, se procede a la estimación de sus parámetros, dado que se trata de un procedimiento iterativo de cálculo, pueden sugerirse valores iniciales.
- g) **Contraste de validez del modelo:** Se utilizan distintos procedimientos para valorar el modelo o modelos inicialmente seleccionados, contraste de significación de parámetros, covarianzas entre estimadores..etc

1. MODELOS LINEALES EN TIEMPO

- h) **Análisis detallado de los errores:** Se tendrán en cuenta las diferencias históricas entre valores reales y estimados por el modelo para su valoración final. Se tiene que verificar el comportamiento no sistemático de los mismos, así como analizar la posible existencia de errores especialmente significativos.
- i) **Selección del modelo:** En base a los resultados de pasos anteriores, se decide sobre el modelo adoptado.
- j) **Predicción:** El modelo seleccionado se utilizara como fórmula inicial de predicción.

1. MODELOS LINEALES EN TIEMPO



2. PROCESO ARIMA

2.1. MODELOS AUTORREGRESIVOS AR (p)

Un modelo autorregresivo **AR** describe una clase particular de proceso en que las observaciones en un momento dado son predecibles a partir de las observaciones previas del proceso mas un termino de error.

El proceso autorregresivo de orden p, representado por ARIMA (p, 0, 0) o simplemente por AR (p):

$$AR(p) = x_t = \varphi_1 x_{t-1} + \varphi_2 x_{t-2} + \cdots + \varphi_p x_{t-p} + a_t$$

Modela nuestra serie en función de los valores rezagados de periodos

2. PROCESO AR₁

2.2. MODELOS DE MEDIAS MÓVILES Ma(q)

Un modelo de medias móviles **MA** describe una serie temporal estacionaria. En este modelo de valor actual pues predecirse a partir de las componente aleatoria de este momento y, en menor medida, de los impulsos aleatorios anteriores.

El proceso de medias móviles de orden q , representado por ARIMA(0,0, q) o también por Ma(q), viene dada por la expresión:

$$MA(q) = x_t = a_t + v_1 a_{t-1} + v_2 a_{t-2} + \cdots + v_q a_{t-q}$$

Modela nuestra serie de tiempo en función de los errores del pasado

2. PROCESO AR₁

Para realizar este tipo de procesos se utiliza el software **STATA 16** (Software estadístico para ciencia de datos).

```
. *Usamos la base de datos de indice de precios al por mayor de EEUU
. use http://www.stata-press.com/data/r15/wpil

. tsset t
    time variable: t, 1960q1 to 1990q4
                delta: 1 quarter

. des wpi
```

variable name	storage type	display format	value label	variable label
wpi	float	%9.0g		U.S. Wholesale Price Index

```
. su wpi
```

Variable	Obs	Mean	Std. Dev.	Min	Max
wpi	124	62.77419	30.24356	30.5	116.2

Propiedades

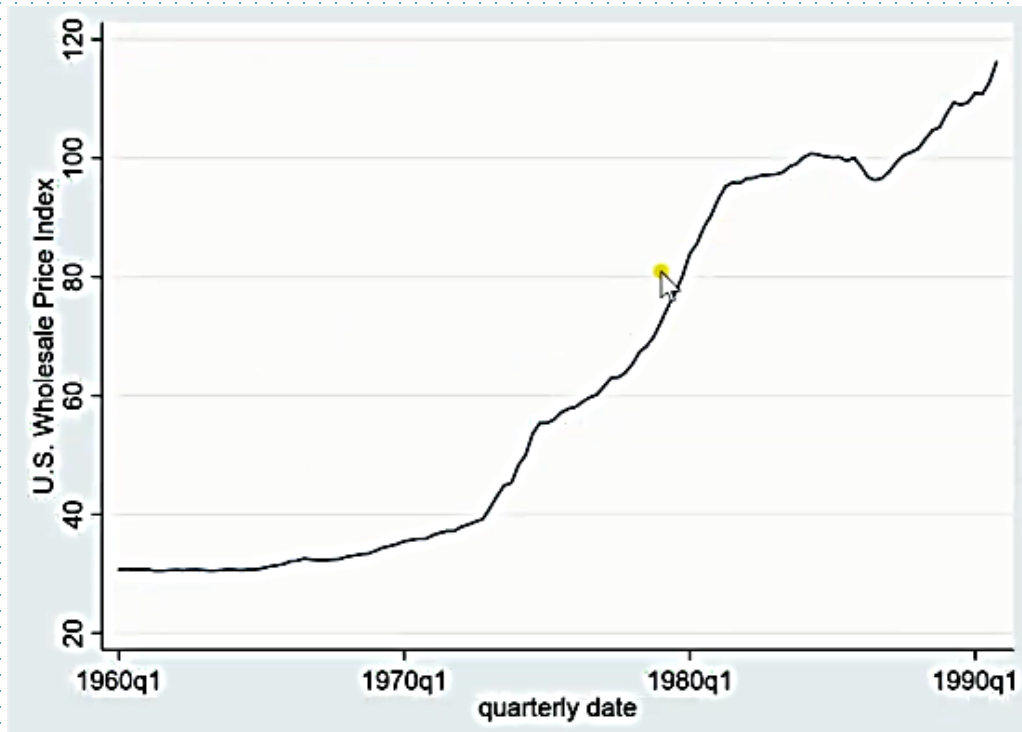
Variables

Nombre	
Etiqueta	
Typo	
Formato	
Etiqueta de valor	
Notas	

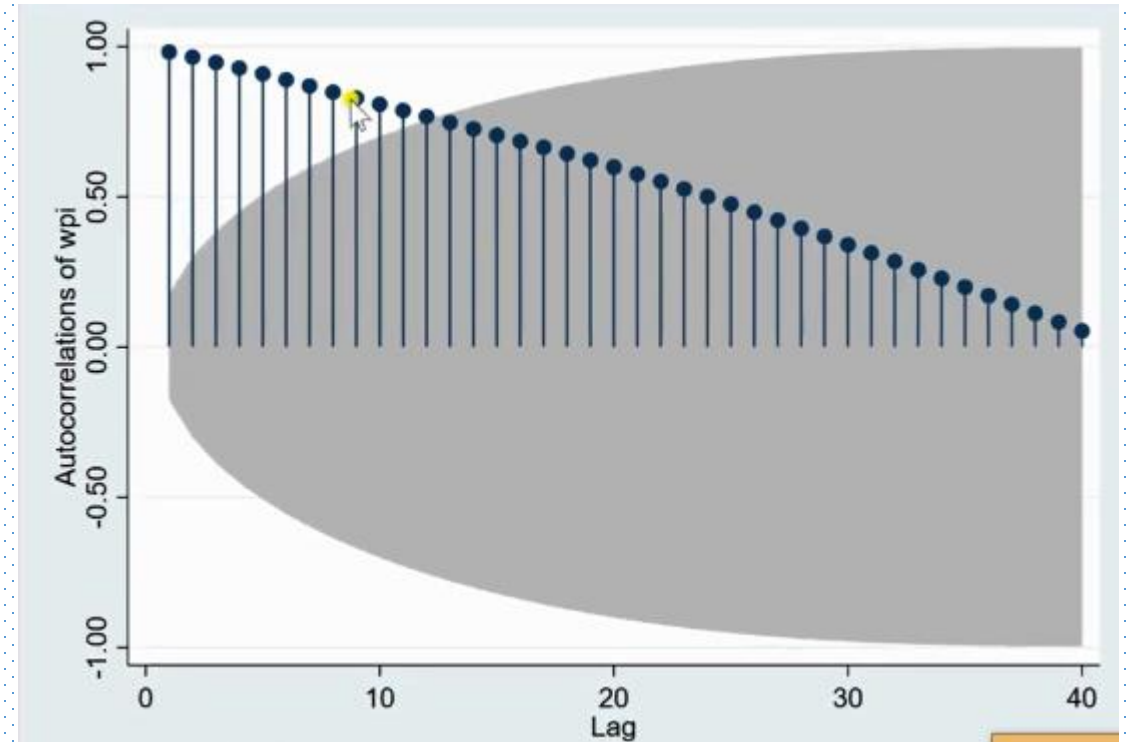
Datos

2. PROCESO AR₁

1er Paso: Analizar la estacionariedad



Tsline



Autocorrelación

2. PROCESO AR₁

Dfuller

```
*
end of do-file

doedit "D:\_
do "C:\User...
do "C:\User...
do "C:\User...
do "C:\User...
do "C:\User...
```

```
*
end of do-file

. do "C:\Users\Edu con Manzanitas\AppData\Local\Temp\STD29b0_000000.tmp"

. tsline wpi, name(gwpi)

*
end of do-file

. do "C:\Users\Edu con Manzanitas\AppData\Local\Temp\STD29b0_000000.tmp"

. ac wpi

*
end of do-file

. do "C:\Users\Edu con Manzanitas\AppData\Local\Temp\STD29b0_000000.tmp"

. dfuller wpi

Dickey-Fuller test for unit root                      Number of obs   =      123

               _____ Interpolated Dickey-Fuller _____
               Test          1% Critical    5% Critical    10% Critical
               Statistic      Value         Value         Value
-----
Z(t)          2.725          -3.502          -2.888          -2.578

MacKinnon approximate p-value for Z(t) = 0.9991

. *HO: la serie es Estacionaria
*
end of do-file
```


2. PROCESO AR₁

2do Paso: Transformación de la serie

```
10
11 *Usamos la base de datos de indice de precios al por mayor de EEUU
12 use http://www.stata-press.com/data/r15/wpi1
13 tsset t
14 des wpi
15 su wpi
16
17 *::::::::::::1 PASO -> ANALIZAR LA ESTACIONARIEDAD ::::::::::::::
18
19 tsline wpi, name(gwpi)
20 *pruebas graficas
21 ac wpi
22
23 *pruebas formales Pruebas de Raiz Unitaria (proximo video)
24 dfuller wpi
25 *H0: la serie es Estacionaria
26
27 *::::::::::::2 PASO -> TRANSFORMACION DE LA SERIE ::::::::::::::
28
29 *si yt no es estacionario-->> diferenciamos
30 gen dwpi= wpi-l1.wpi
31 gen dwpi2= d.ln_wpi
32
33 tsline dwpi, name(gdwpi)
34 graph combine gwpi gdwpi
35 dfuller dwpi
36
37 *::::::::::::3 PASO -> IDENTIFICAR EL MODELO ::::::::::::::
38 ac dwpi, name(gac)
```

Nombre	Etiqueta
wpi	U.S. Wholesale Pri...
t	quarterly date
ln_wpi	
dwpi	
dwpi2	

Propiedades

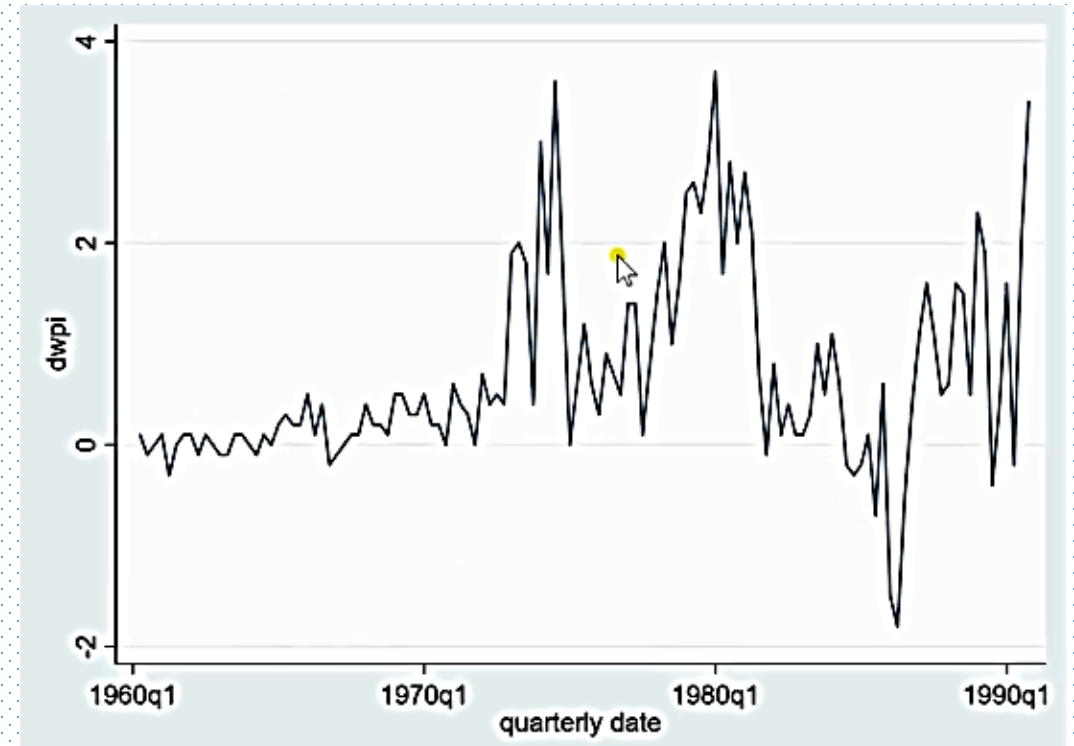
Variables

Nombre	Etiqueta	Typo	Formato	Etiqueta de valor	Notas
dwpi					

Datos

Nombre de archivo	Etiqueta	Notas	Variables	Observaciones	Tamaño
wpi1.dta					2.18K

Linea: 31, Col: 20 CAP. NUM. OVR



Autocorrelación transformada

2. PROCESO AR₁

Dfuller
transformada

```
. do "C:\Users\Edu con Manzanitas\AppData\Local\Temp\STD29b0_000000.tmp"
. gen dwpi= wpi-l1.wpi
(1 missing value generated)
. gen dwpi2= d.ln_wpi
(1 missing value generated)
.
end of do-file

. do "C:\Users\Edu con Manzanitas\AppData\Local\Temp\STD29b0_000000.tmp"
. tset dwpi, name(gdwpi)
.
end of do-file

. do "C:\Users\Edu con Manzanitas\AppData\Local\Temp\STD29b0_000000.tmp"
. dfuller dwpi
Dickey-Fuller test for unit root          Number of obs   =      122

              Test              _____ Interpolated Dickey-Fuller _____
              Statistic          1% Critical   5% Critical   10% Critical
                                Value          Value          Value
-----
Z(t)          -4.661            -3.503        -2.889        -2.579

MacKinnon approximate p-value for Z(t) = 0.0001
.
end of do-file
```

Comando

wpi	U.S. Wholesale Pri...
t	quarterly date
ln_wpi	
dwpi	
dwpi2	


Propiedades

Variables

Nombre	
Etiqueta	
Typo	
Formato	
Etiqueta de valor	
Notas	

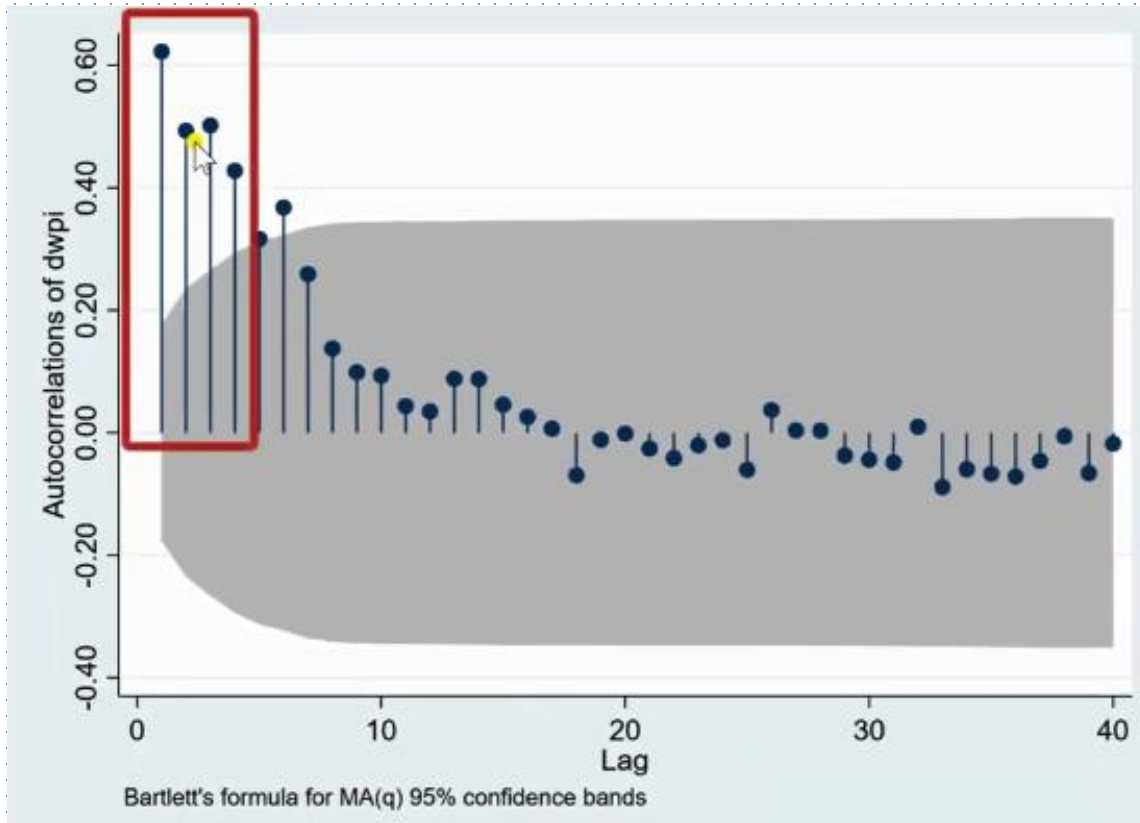
Datos

Nombre de archivo	wpi1.dta
Etiqueta	
Notas	
Variables	5
Observaciones	124
Tamaño	2.18K
Memoria	64M
Ordenado por	

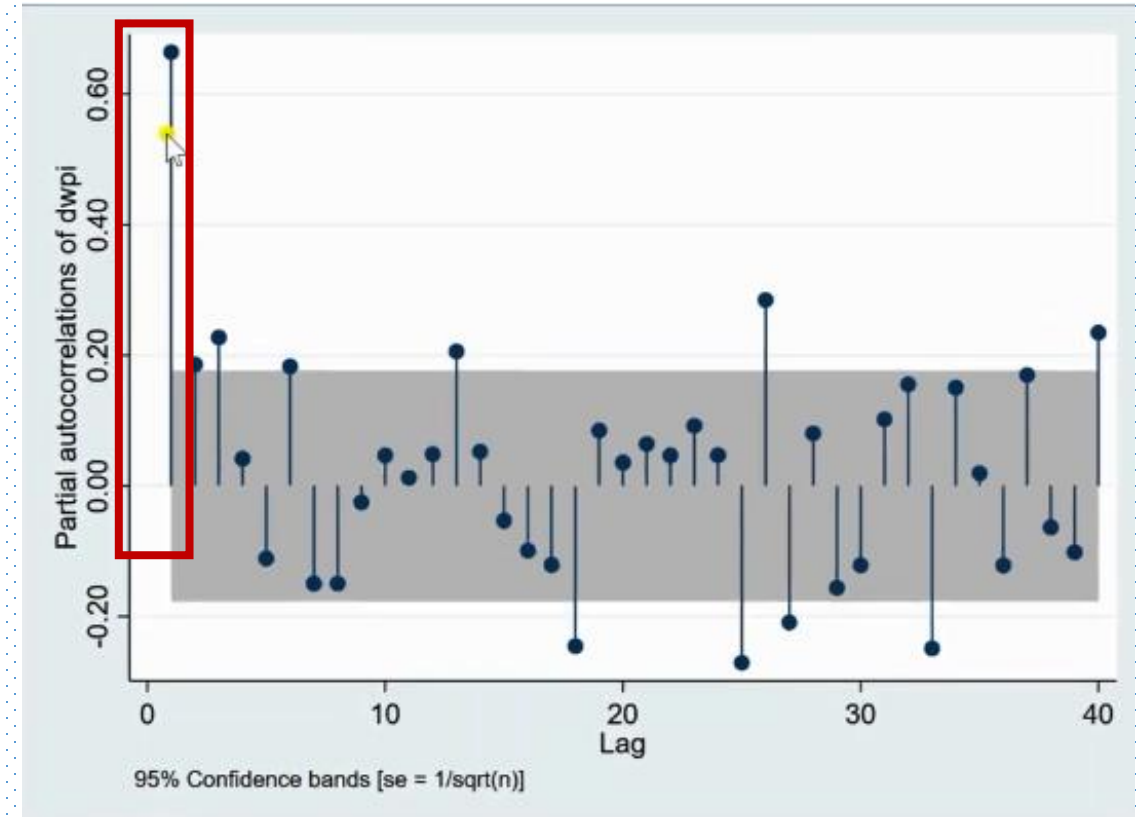
 Suscribirse

2. PROCESO AR₁

3er Paso: Identificación del modelo



Medias móviles



Autoregresión

3. RUIDO BLANCO

El ruido blanco o sonido blanco es una **señal aleatoria** (proceso estocástico) que se caracteriza por el hecho de que sus valores de señal en dos tiempos diferentes **no guardan correlación estadística**.

Como consecuencia de ello, su densidad espectral de potencia (DSP) es una constante, es decir, su gráfica es plana.

Esto significa que la señal contiene todas las frecuencias y todas ellas muestran la misma potencia. Igual fenómeno ocurre con la luz blanca, de allí la denominación.



3. RUIDO BLANCO

El ruido blanco es una señal no correlativa, es decir, en el eje del tiempo la señal toma valores sin ninguna relación unos con otros.

Cuando se dice que tiene una densidad espectral de potencia plana, con un ancho de banda teóricamente infinito, es que en un gráfica espectral de frecuencia tras haber realizado una descomposición espectral de Fourier, en el dominio de la frecuencia veríamos todas los componentes con la misma amplitud, haciendo el efecto de una línea continua paralela al eje horizontal.



3. RUIDO BLANCO



4to Paso: Validación del modelo

```
11 arima-box jenkins.do
12 *AC -> orden de los MA(4)
13 *PAC-> orden de los AR(1)
14
15 arima dwpi, arima(1,0,1)
16 arima dwpi, arima(1,0,2)
17 arima dwpi, arima(1,0,3)
18 arima dwpi, arima(1,0,4)
19
20 *:::::::::: 4 PASO -> VALIDACION DEL MODELO ::::::::::::::
21
22 *mejor modelo es el mas parsimonioso
23
24 *test de portmanteau
25 *H0: el proceso es ruido blanco
26 *---ARIMA(1,1)
27 quietly arima dwpi, arima(1,0,1)
28 predict resid11, resid
29 wntestq resid11
30 wntestb resid11
31
32 estat ic
33 *la media del error debe ser cero
34 summarize resid11
35 scalar media_error11= r(mean)
36 tsline resid11, yline(`r(mean)')
37
38 *---ARIMA(1,2)
39 quietly arima dwpi, arima(1,0,2)
```

Linea: 52, Col: 39 CAP NUM OVR

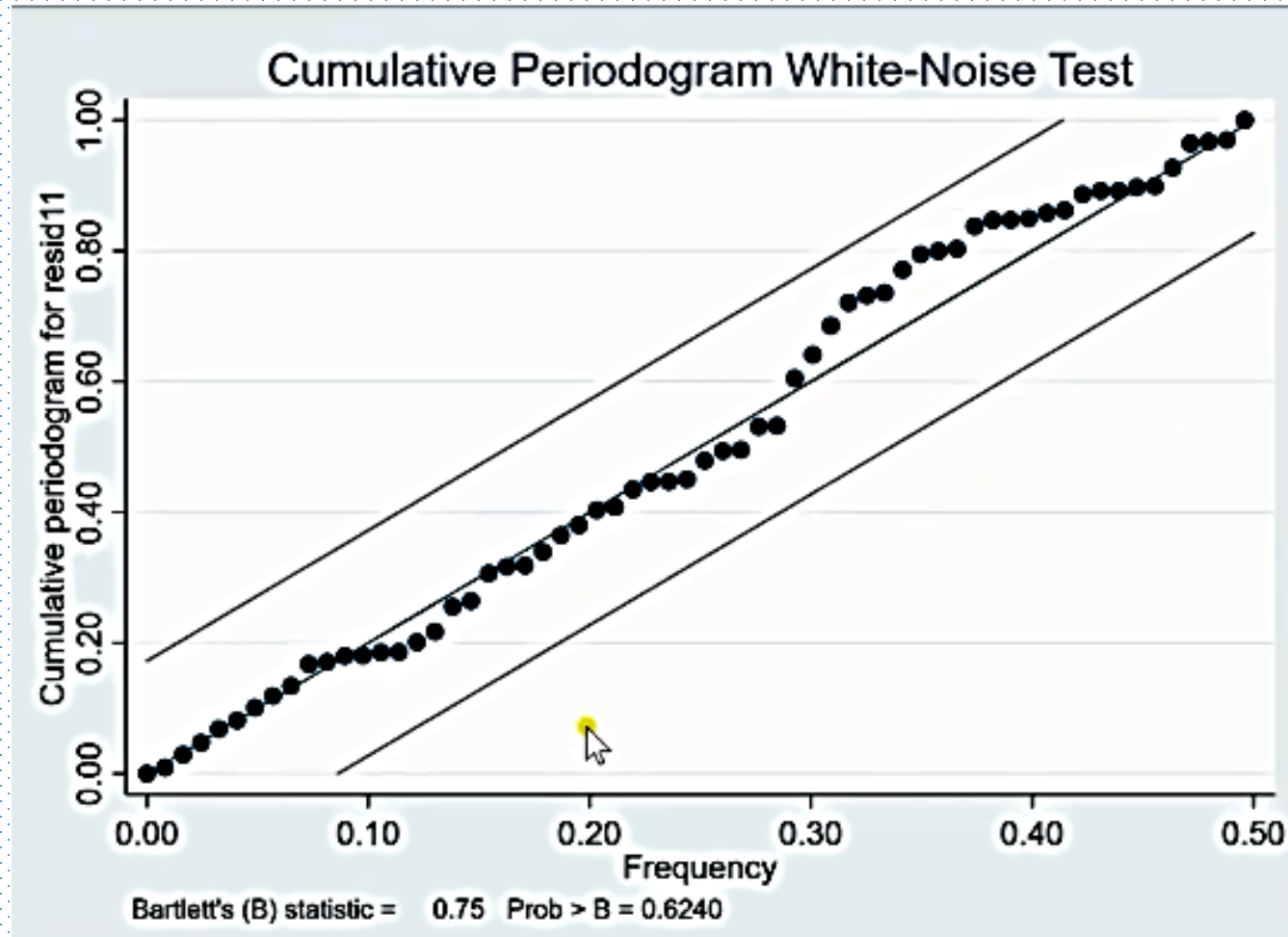
dwpi	_cons	.7472247	.3526501	2.12	0.034	.0560433	1.438406
ARMA							
ar	L1.	.8910989	.0752727	11.84	0.000	.7435671	1.038631
	L2.	-.3867376	.1156022	-3.35	0.001	-.6133136	-.1601615
ma	L1.	-.0827953	.0898423	-0.92	0.357	-.258883	.0932924
	L2.						
/sigma		.7233135	.0369548	19.57	0.000	.6508935	.7957434

Note: The test of the variance against zero is one sided, and the two-sided confidence interval is truncated at zero.

```
end of do-file
do "C:\Users\Edu con Manzanas\AppData\Local\Temp\STD29b0_000000.tmp"
quietly arima dwpi, arima(1,0,1)
predict resid11, resid
(1 missing value generated)
wntestq resid11
Portmanteau test for white noise
Portmanteau (Q) statistic = 36.9722
Prob > chi2(40) = 0.6073
end of do-file
```

Validación del ruido blanco por cada modelo

3. RUIDO BLANCO



4. PROCESOS GAUSSIANNOS

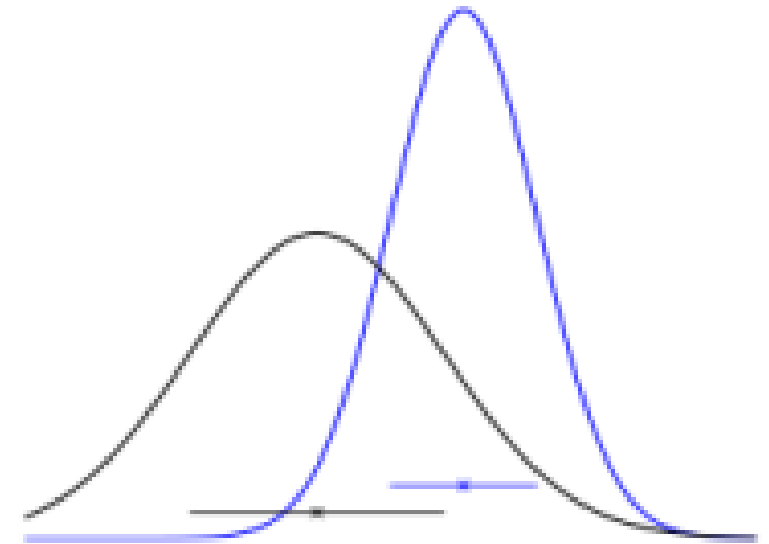


- Normalmente, en los algoritmos de aprendizaje que hemos visto, dado un conjunto de ejemplos de entrenamiento se busca encontrar el mejor modelo que ajuste a los datos (o el modelo que haga las mejores predicciones con ejemplos de prueba).
- Cuando hablamos de algoritmos gaussianos, podemos buscar la distribución posterior sobre los modelos para su solución.
- Estas distribuciones cuantifican nuestra incertidumbre en los modelos.

4. PROCESOS GAUSSIANOS

4.1. DEFINICIÓN

- Los procesos gaussianos se definen como una distribución de probabilidad sobre funciones aleatorias.
- De hecho son sobre colecciones infinitas de variables (funciones), tal que cualquier subconjunto de variables aleatoria finita tiene una distribución gaussiana multivariable.



4. PROCESOS GAUSSIANOS

La distribución de GAUSS o distribución normal es fundamental en la aplicación de la inferencia estadística, ya que las distribuciones de muchos estadígrafos muestrales tienden a la distribución Normal conforme crece el tamaño de la muestra.

Se dice que una variable aleatoria **X** está normalmente distribuida si su función de densidad está dada por:

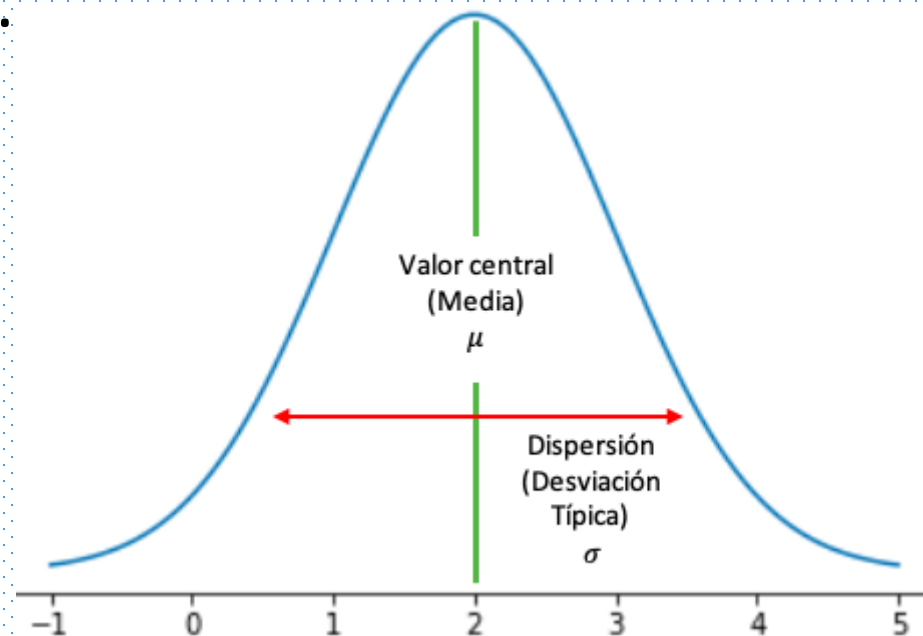
$$f(x) = \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2\pi\sigma^2}}$$

$$-\infty < x < +\infty$$

4. PROCESOS GAUSSIANOS

i. Gráfica de la Distribución Normal

La gráfica de la distribución Normal es una curva simétrica con forma de campana, que se extiende sin límites tanto en la dirección positiva como en la negativa.



4. PROCESOS GAUSSIANOS

ii. Cálculo de probabilidades

La probabilidad de que una variable aleatoria normalmente distribuida sea menor o igual a un valor específico, esta dada por la función de distribución acumulada de la siguiente ecuación:

$$p(x \leq a) = \int_{-\infty}^a \frac{e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}}{\sqrt{2\pi\sigma^2}}$$

Por tanto, para reducir el problema anterior, es necesario standarizar la variable, de tal forma que permita presentar los resultados en una sola tabla, es decir:

4. PROCESOS GAUSSIANOS

$$z = \frac{x - \mu}{\sigma}$$

Luego:

$$f(z) = \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi\sigma^2}}$$

Con

$$E(z) = 0$$

$$V(z) = 1$$

De manera que:

$$p(x \leq a) = p\left(z \leq \frac{a - \mu}{\sigma}\right) = \int_{-\infty}^{\frac{a - \mu}{\sigma}} \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} dz$$

Tabla Normal
Estandarizado

4. PROCESOS GAUSSIANOS

iii. Propiedad reproductiva de la Distribución Normal

Una propiedad muy importante de la Distribución Normal es la llamada propiedad Reproductiva o Aditiva de la distribución normal, la cual indica:

Pob 1 (x_1)
 $\mu_1; \sigma_1^2$

Pob 2 (x_2)
 $\mu_2; \sigma_2^2$

Pob K (x_k)
 $\mu_k; \sigma_k^2$

Entonces se cumple: $y = ax_1 + bx_2 + cx_3 + \dots + kx_k$

$$E(y) = \mu_y = a\mu_1 + b\mu_2 + c\mu_3 + \dots + k\mu_k$$

$$V(y) = \sigma_y^2 = a\sigma_1^2 + b\sigma_2^2 + c\sigma_3^2 + \dots + k\sigma_k^2$$

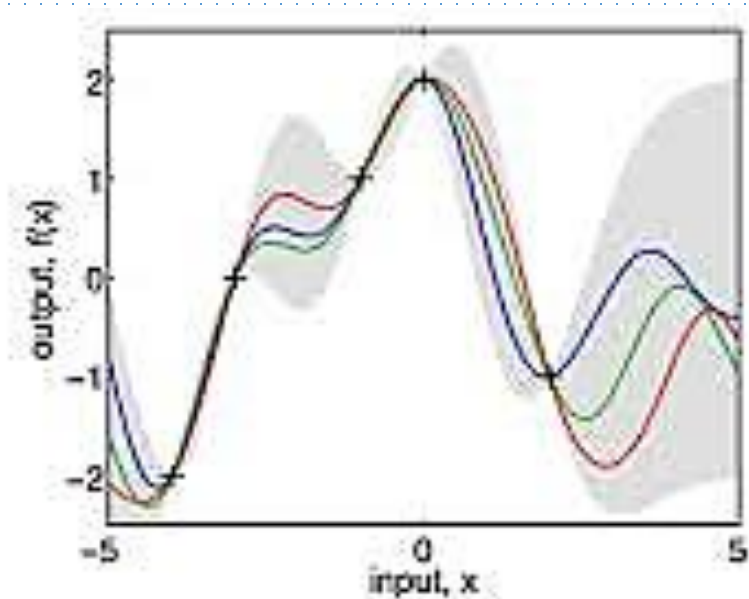
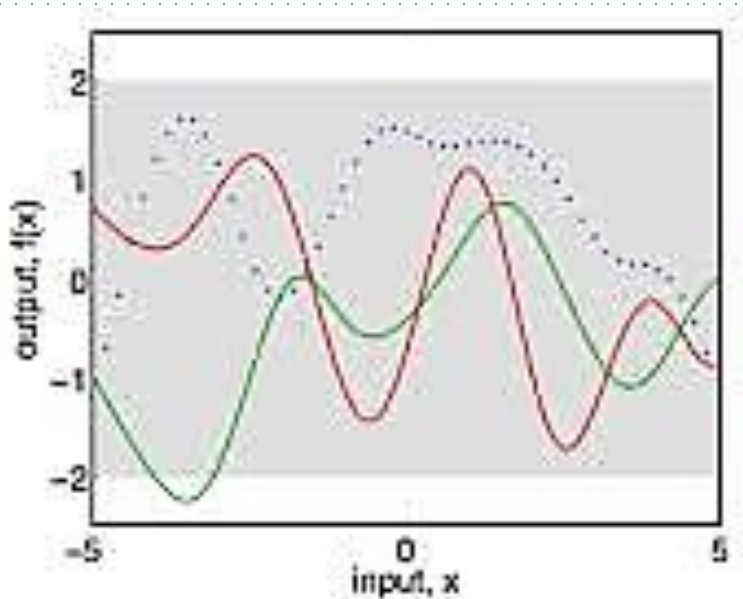
4. PROCESOS GAUSSIANOS

4.2. PROCESOS GAUSSIANOS

- Un proceso gaussiano es un proceso estocástico tal que cualquier subconjunto de variables aleatoria finita tiene una distribución gaussiana multivariable.
- Los procesos gaussianos son una extensión hacia colecciones infinitas de variables.
- Esta extensión nos permite pensar en los procesos gaussianos como distribuciones, no solo sobre vectores aleatorios sino sobre distribuciones de funciones aleatorias.

4. PROCESOS GAUSSIANOS

- En particular, una colección de variables aleatorias $\{h(x) : x \in X\}$ se obtiene de un proceso gaussiano con una función media (μ) y una función de covarianza (COV), si para cualquier conjunto finito de elementos $x_1, \dots, x_m \in X$, el conjunto finito de variables aleatorias asociadas $h(x_1), \dots, h(x_m)$ tienen la siguiente distribución.



5. PRUEBAS NO PARAMÉTRICAS

5.1. INTRODUCCIÓN

Hasta ahora la mayor parte de las pruebas estadísticas (pruebas de hipótesis estadísticas, análisis de varianza, ajuste de curvas de regresión) e intervalos de confianza se basan en ciertos supuestos, por lo cual, han sido denominados métodos paramétricos.

- Las **pruebas paramétricas** se basan en el análisis de un parámetro poblacional cuyo estimador tiene una distribución conocida (generalmente una distribución normal) o puede aproximarse a una distribución normal.
- Pero en el caso de no cumplirse alguno de estos supuestos es necesaria la aplicación de las denominadas **pruebas no paramétricas** o pruebas de distribución libre.

5. PRUEBAS NO PARAMÉTRICAS

Las **pruebas no paramétricas** se utilizan:

- Cuando se tiene duda de que las observaciones pertenecen a una distribución normal.
- Cuando se tienen muestras pequeñas con distribuciones desconocidas (por lo que ya no es posible aplicar el TCL)
- Para probar hipótesis sobre la forma y posición de las distribuciones

➤ **Ventajas**

- Se pueden aplicar a datos de tipo cuantitativo y cualitativo
- Son rápidas y fáciles de realizar

5. PRUEBAS NO PARAMÉTRICAS

➤ Desventajas

- En el caso de tener la posibilidad de realizar pruebas paramétricas y no paramétricas para una determinada situación, es mejor efectuar las pruebas paramétricas por ser mas precisas.
- Las pruebas no paramétricas son menos eficientes puesto que no utilizan toda la información proveniente de la muestra (lo cual implica incrementar el tamaño de la muestra).



5. PRUEBAS NO PARAMÉTRICAS

5.2. PRUEBA DE CORRIDAS PARA LA ALEATORIEDAD

En todas las pruebas anteriores uno de los supuestos fundamentales era el hecho de la existencia de aleatoriedad en la toma de datos. En la practica, no siempre es posible controlar la forma en la que han sido tomados los datos, por lo que, es necesario efectuar una prueba para determinar la existencia o no de aleatoriedad en los datos.

La prueba de corridas se utiliza:

- Para analizar la existencia de aleatoriedad en los datos recolectados, considerando el orden en el que han sido obtenidos.
- Para determinar si existe alguna tendencia en los datos.

5. PRUEBAS NO PARAMÉTRICAS

Una corrida se define al conjunto de observaciones similares contenidas dentro de un conjunto de observaciones diferentes

En esta prueba se puede presentar dos casos:

- **Datos cualitativos:** Para lo cual los datos se dividen en dos categorías
- **Datos cuantitativos:** Para lo cual los datos se dividen en dos categorías, en función a si están por encima o por debajo de la mediana de la muestra.

5. PRUEBAS NO PARAMÉTRICAS

Los pasos a seguir son:

1. Plantear las hipótesis

H_0 : Las observaciones han sido recolectados en forma aleatoria

H_1 : Las observaciones no han sido recolectadas en forma aleatoria

2. Especificar α %

3. Determinación del estadígrafo

Para efectuar la prueba se determina el estadígrafo R que es el numero de corridas en la muestra.

5. PRUEBAS NO PARAMÉTRICAS

4. Establecer la región de aceptación. Para este caso se define lo siguiente:

R = Número de corridas en la muestra

n1 = Número de elementos en la muestra del primer tipo

n2 = Número de elementos en la muestra del segundo tipo

- Si **n1 ≤ 10** o **n2 ≤ 10** se emplea la prueba C (con tablas específicas para tal efecto).
- Si **n1 > 10** y **n2 > 10** se puede aproximar a la distribución normal previa estandarización de R y por lo cual la región de aceptación es:

$$\left(-Z_{1-\frac{\alpha}{2}} ; +Z_{1-\frac{\alpha}{2}} \right)$$

5. PRUEBAS NO PARAMÉTRICAS

5. Calcular el valor de z empleando una muestra de tamaño n y estandarizar.

$$Z = \frac{(R - \mu_R)}{\sigma_R}$$

En la que:

$$\mu_R = \frac{2 n_1 n_2}{n_1 + n_2} + 1$$

$$\sigma_R = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2 * (n_1 + n_2 - 1)}}$$

5. PRUEBAS NO PARAMÉTRICAS

6. Tomar la decisión

Si $Z \in \alpha$:

$$\left(-z_{1-\frac{\alpha}{2}} ; +z_{1-\frac{\alpha}{2}}\right)$$

Entonces:

Se acepta H_0

Caso contrario:

Se rechaza H_0



5. PRUEBAS NO PARAMÉTRICAS

1. Determina el numero de corridas, según los siguientes datos:

M F F F M M F M F F M M

2. Determine el numero de corridas de las siguientes edades:

17; 6; 9; 18; 25; 11; 7

Criterios para la prueba:

Cuando el numero de corridas es pequeño o grande al número de muestras, no existe aleatoriedad.

5. PRUEBAS NO PARAMÉTRICAS

5.3. PRUEBA SOBRE LA MEDIANA DE UNA POBLACIÓN

Para probar la hipótesis respecto al valor de la mediana de una población se tiene la prueba denominada prueba del **rango con signo** o de **Wilcoxon**. Esta prueba es la equivalente a la prueba paramétrica sobre el valor de la media poblacional.

Se requiere que los valores de la muestra aleatoria se encuentren por lo menos en una escala ordinal y no es necesario efectuar suposiciones respecto a la forma de la distribución de probabilidad.

Puesto que la distribución normal es simétrica, la media aritmética de una distribución normal es igual a la mediana, por lo cual, la prueba Wilcoxon puede emplearse para probar hipótesis respecto a la media aritmética de una distribución Normal.

5. PRUEBAS NO PARAMÉTRICAS

La hipótesis nula y alternativa se plantean con respecto a la mediana poblacional y puede ser una o de dos colas. Para esta prueba se utilizan tablas especiales, llamadas tablas de Wilcoxon.

Los pasos a seguir (para el caso bilateral) son:

1. $H_0: Me = a$
 $H_1: Me \neq a$
2. Especificar $\alpha \%$
3. Se determina el estadígrafo W
4. Definir el criterio de aceptación
5. Calcular W
6. Tomar decisión

5. PRUEBAS NO PARAMÉTRICAS

5.4. PRUEBA DE PEARSON

Esta prueba se utiliza con preferencia en el caso de que los datos:

- a) Se ajustan a una distribución de variable aleatoria discreta
- b) El tamaño de muestra es grande

Para aceptar la hipótesis nula, debe ser posible atribuir las diferencias entre las frecuencias observadas y las esperadas a la variabilidad del muestreo y al nivel de significancia.

Es así que el estadígrafo de prueba de Pearson se basa en la magnitud de esta diferencia para cada una de las categorías de la distribución de frecuencias.

5. PRUEBAS NO PARAMÉTRICAS

El valor de dicho estadígrafo es:

$$\pi = \sum_{i=1}^k \frac{(foi - foe)^2}{fei}$$

En la que:

- Foi= frecuencia observada absoluta
- Foe= frecuencia esperada absoluta

Por otro lado es posible demostrar que el estadígrafo π sigue una distribución Chi cuadrado, por lo que se aplica dicha distribución y es por ese motivo que también recibe ese nombre .

5. PRUEBAS NO PARAMÉTRICAS

Los pasos a seguir son:

1. Establecer la hipótesis

Ho: las frecuencias están distribuidas según una distribución determinada

H1: las frecuencias no están distribuidas según una distribución determinada

2. Especificar el valor de α %

3. Se debe utilizar el estadístico π

4. Definir la región de aceptación

5. Calcular el valor de π

6. Tomar la decisión

6. ECUACIÓN DE CHAPMAN-KOLMOGOROV - SMIRNOV

La prueba de Kolmogorov – Smirnov es la prueba de bondad de ajuste que se aplica en casos en que:

- a) Los datos provienen de una distribución continua de probabilidades
- b) El tamaño de muestra es pequeño

Esta prueba se basa en una comparación entre los valores de la distribución acumulada que se observan en la muestra respecto a los valores de la distribución acumulada de la distribución propuesta en la hipótesis nula.

Si esta comparación revela una diferencia suficientemente grande entre dichos valores, entonces la hipótesis nula se rechaza.

6. ECUACIÓN DE CHAPMAN-KOLMOGOROV

Lo anterior se evalúa con la ecuación:

$$D = \text{Max } |FOi - FEi|$$

En la que:

Foi= Frecuencia observada acumulada relativa de la observación “i”

Fei= Frecuencia esperada acumulada relativa de la observación “i”

En este tipo de prueba se utiliza una tabla especial llamada Tabla de Kolmogorov – Smirnov.

6. ECUACIÓN DE CHAPMAN-KOLMOGOROV

Los pasos a seguir son:

1. Establecer las hipótesis

H0: Las frecuencias están distribuidas según una distribución determinada

H1: Las frecuencias no están distribuidas según una distribución determinada

2. Especificar el valor de α %

3. Se debe utilizar el estadístico D

4. Definir la región de aceptación

5. Calcular el valor D, ordenando previamente los datos en forma ascendente

6. Tomar la decisión