



**Universidad De Carabobo**  
**Facultad De Ingeniería**  
**Escuela De Ingeniería Industrial**



**DESARROLLO DE UNA METODOLOGÍA COMPUTARIZADA PARA EL ANÁLISIS  
DE DATOS PROVENIENTES DE PROCESOS PRODUCTIVOS Y LA  
CONSTRUCCIÓN DE LOS GRÁFICOS DE CONTROL EWMA, CUSUM Y  
ANÁLISIS DE PUNTO DE CAMBIO.**

**Tutor Académico:**

Dr. Teodoro García.

Br. Javier Gómez

**Integrantes:**

Br. Carolina Pérez

Valencia, Noviembre de 2012.

## CAPÍTULO I



## CAPÍTULO II

## CAPÍTULO III

## CAPÍTULO IV

## CAPÍTULO V





## CONCLUSIONES

## RECOMENDACIONES

## APÉNDICES

## BIBLIOGRAFÍA





**Universidad De Carabobo  
Facultad De Ingeniería  
Escuela De Ingeniería Industrial**



**DESARROLLO DE UNA METODOLOGÍA COMPUTARIZADA PARA EL ANÁLISIS DE  
DATOS PROVENIENTES DE PROCESOS PRODUCTIVOS Y LA CONSTRUCCIÓN  
DE LOS GRÁFICOS DE CONTROL EWMA, CUSUM Y ANÁLISIS DE PUNTO DE  
CAMBIO.**

Trabajo Especial de Grado presentado ante la Ilustre Universidad de Carabobo, para optar al  
Título de Ingeniero Industrial

Línea de investigación: Métodos Estadísticos Multivariantes.

**Tutor Académico:**

Dr. Teodoro García  
Br. Carolina Pérez

**Autor:**

Br. Javier Gómez

Valencia, Noviembre de 2012.



**Universidad De Carabobo**  
**Facultad De Ingeniería**  
**Escuela De Ingeniería Industrial**



**CERTIFICADO DE APROBACIÓN**

Quienes suscriben, Miembros del Jurado designado por el Consejo de Escuela de la Facultad de Ingeniería de la Universidad de Carabobo, para examinar el Trabajo Especial de Grado titulado **“Desarrollo de una metodología computarizada para el análisis de datos provenientes de procesos productivos y la construcción de los gráficos de control EWMA, CUSUM y Análisis de Punto de Cambio”**, el cual está adscrito a la Línea de Investigación **“Métodos Estadísticos Multivariantes”** del Departamento de **“Investigación de Operaciones”**, presentado por los Bachilleres **Carolina Pérez, C.I. 15.365.961** y **Javier Gómez, C.I. 17.025.655**, a los fines de cumplir con el requisito académico exigido para optar al Título de Ingeniero Industrial, dejan constancia de lo siguiente:

1. Leído como fue dicho Trabajo Especial de Grado, por cada uno de los Miembros del Jurado, éste fijó el día Viernes 23 de Noviembre de 2012, a las 2:00 pm, para que los autores lo defendieran en forma pública, lo que éstos hicieron, en el Salón SDC, mediante un resumen oral de su contenido, luego de lo cual respondieron satisfactoriamente a las preguntas que le fueron formuladas por el Jurado, todo ello conforme a lo dispuesto en el

Reglamento del Trabajo Especial de Grado de la Universidad de Carabobo y a las Normas de elaboración de Trabajo Especial de Grado de la Facultad de Ingeniería de la misma Universidad.

2. Finalizada la defensa pública del Trabajo Especial de Grado, el Jurado decidió aprobarlo por considerar que se ajusta a lo dispuesto y exigido por el Reglamento de Estudios de Pregrado.

En fe de lo cual se levanta la presente acta, a los 23 días del mes de Noviembre de 2012, dejándose también constancia de que actuó como Coordinador del Jurado el Tutor, Prof. Teodoro García.

Firma del Jurado Examinador

Prof. Teodoro García  
Tutor

Prof. Manuel Jiménez  
Prof. Carlos Martínez  
Jurado

Jurado





---

## DEDICATORIA

A Dios.

A mi hermano.

A mis padres.

A mis profesores.

A mis compañeros de la universidad.

A mis amigos y familiares.

*Carolina Pérez*



A Dios y a La Virgen María por darme la perseverancia e inteligencia necesaria para alcanzar esta meta, por ayudarme a vencer los obstáculos que se me presentaron a lo largo de este camino y estar siempre a mi lado.

A mis padres, Eduardo y Norelys por ser el pilar fundamental en todo lo que soy, en toda mi educación, tanto académica, como de la vida, por su incondicional apoyo perfectamente mantenido a través del tiempo.

A mis hermanos Kalem, Eduardo, John y todo el resto de mi familia por apoyarme a su manera en seguir hacia adelante.

A mis profesores, quienes han sido parte esencial a través de su aporte en la formación integral y profesional.

A mis compañeros de estudios, quienes han estado en todo momento compartiendo y apoyándome en este largo camino. Gracias por estar presentes.



## Dedicatoria

---



*Mil bendiciones a todos.*

*Javier Gómez.*



---

## AGRADECIMIENTOS

A mi hermano por ser el cimiento de mi vida y creer siempre en mí.

A mi tutor de tesis: Dr. Teodoro García por guiarme y motivarme siempre a seguir en este proceso de aprendizaje.

A mis padres por el apoyo prestado durante toda mi carrera.

A mis familiares (Nena, Estela, Aníbal, Gustavo) por ser ejemplo de constancia y superación.

A mis compañeros de la universidad (Gabriel, Misle, Henry, Fran, Ruben, Blanquito, Ronald, Javier) que siempre me acompañan, ayudan y me impulsan a ser y dar lo mejor de mí.

A todas aquellas personas que de una u otra forma dieron su aporte para la realización de este trabajo especial de grado.

A la Universidad de Carabobo por todo lo que me dio.

*Carolina Pérez*



Al Ingeniero por excelencia, diseñador y constructor de la vida; sin su presencia, dirección y sabiduría sería imposible la materialización de este triunfo.

A mis padres, mis hermanos y demás familiares por el apoyo prestado durante toda mi carrera.

Al Dr. Teodoro García, que con sus constantes aportes se logró la conclusión de éste trabajo.

A todos mis amigos y compañeros de la universidad que estuvieron a mi lado en esta etapa de mi vida.

A mi compañera de tesis Carolina, que por vueltas del destino nos emparejó para realizar este proyecto.



## Agradecimientos

---



A la ilustre Universidad de Carabobo, generadora de conocimientos y punto de partida para que se cumpliera el sueño de ser hoy Ingeniero Industrial.

*Javier Gómez*



---

**UNIVERSIDAD DE CARABOBO**  
**FACULTAD DE INGENIERÍA**  
**ESCUELA DE INGENIERÍA INDUSTRIAL**

**DESARROLLO DE UNA METODOLOGÍA COMPUTARIZADA PARA EL ANÁLISIS  
DE DATOS PROVENIENTES DE PROCESOS PRODUCTIVOS Y LA  
CONSTRUCCIÓN DE LOS GRÁFICOS DE CONTROL EWMA, CUSUM Y  
ANÁLISIS DE PUNTO DE CAMBIO.**

Autores: Pérez Carolina

Gómez Javier

Tutor: Dr. García Teodoro

Fecha: Noviembre de 2012.

**RESUMEN**

En el presente trabajo especial de grado, se desarrolla una metodología computarizada para analizar los datos provenientes de procesos productivos y establecer las medidas remediales en caso de ser necesarias con el objeto de cumplir con los supuestos teóricos para el diseño de los gráficos de control CUSUM, EWMA y Análisis de Punto de Cambio. La metodología combina el análisis exploratorio con el análisis inferencial de datos permitiendo reconocer la presencia de múltiples outliers y el cumplimiento de supuestos tales como la normalidad, estacionariedad y ausencia de autocorrelación en los datos, pudiéndose aplicar medidas remediales. La herramienta permite la construcción de los gráficos de control antes mencionados para datos independientes y autocorrelacionados. Para validar la metodología computarizada se utilizan datos teóricos y reales provenientes de la industria papelera para demostrar la forma de llevar a cabo un control estadístico de calidad en un proceso industrial.



---

**Palabras Clave:** Análisis de datos, CUSUM, EWMA, Análisis de Punto de Cambio, Estacionariedad, Autocorrelación.





## ÍNDICE DE FIGURAS

FIGURA 2.1. Representación gráfica del diagrama de caja. ....	22
FIGURA 2.2. Regiones del estadístico de Durbin y Watson. ....	51
FIGURA 2.3. Valores de $c$ para diferentes valores de $\lambda, \varphi$ y $\Psi$ . .....	81
FIGURA 2.4. Valores de $c$ para diferentes valores de $r, \varphi, \psi$ . .....	84
FIGURA 3.1. Procedimiento seguido en la investigación. .....	92
FIGURA 3.2. Diagrama de la fase de Análisis Exploratorio de los datos. .....	94
FIGURA 3.3. Diagrama de flujo del estudio de estacionariedad. .....	96
FIGURA 3.4. Diagrama de flujo del estudio de autocorrelación. .....	98
FIGURA 3.5. Diagrama de flujo de la metodología Box-Jenkins. .....	100
FIGURA 3.6. Diagrama de flujo del gráfico de control EWMA. .....	102



---

FIGURA 3.7. Diagrama de flujo del gráfico de control CUSUM. .....	103
FIGURA 3.8. Diagrama de flujo del gráfico de control EWMA autocorrelacionado. .....	104
FIGURA 3.9. Diagrama de flujo del gráfico de control CUSUM autocorrelacionados. .....	106
FIGURA 3.10. Diagrama de flujo del Análisis de Punto de Cambio. .....	108
FIGURA 4.1. Diagrama de flujo General del programa. ....	112



## ÍNDICE DE TABLAS

TABLA A.1. Correspondiente al estadístico $d$ de Durbin - Watson: puntos significativos de DU al nivel de significancia de 0.05 .....	235
TABLA A.2 Correspondiente a la elección de los valores de $L$ y $K$ en la construcción del grafico EWMA para datos independientes.....	236
TABLA A.3. Correspondiente a un $\psi=0.5$ para la elección de los valores de $\lambda$ en la construcción de grafico EWMA para datos Autocorrelacionados.....	238
TABLA A.4. Correspondiente a la obtención del valor de $c$ en el gráfico EWMA Autocorrelacionados.....	242
TABLA A.5. Correspondiente a la elección de los valores de $r$ en gráficos de CUSUM autocorrelacionados.....	244
TABLA A.6. Correspondiente a la elección de $C$ en la obtención del gráfico CUSUM autocorrelacionados.....	24



---

TABLA A.7. Límite de Control  $h_{n,\alpha}$  para una muestra de Tamaño  $n$  , rechazo  $\alpha$ , a partir de  $n = 10$ . ..... 248



## INTRODUCCIÓN

En la industria existen situaciones donde es necesario el control o monitoreo simultáneo de una o más variables de calidad del proceso productivo. El control estadístico de procesos (SPC) es una colección de herramientas – gráficos de control – de resolución de problemas, útil para lograr la estabilidad y mejorar la capacidad del proceso mediante la reducción de la variabilidad. El objetivo del SPC es lograr la calidad del producto, reducción de costos de producción, minimizando los defectos del producto.

Cuando el interés se centra en la detección de cambios en la media y/o varianza del proceso, los cuales sean menores a dos veces sigma, los gráficos de control CUSUM y EWMA representan una buena opción, pues han demostrado tener gran capacidad para detectar dichos cambios. Sin embargo, dado que los gráficos de control no muestran información detallada del cambio ocurrido, es decir, solo informan sobre los parámetros de control del proceso, se ha desarrollado para mejorar esas deficiencias, el Análisis de Punto de Cambio, la cual es una herramienta estadística muy eficaz y poderosa en la detección de pequeños cambios sostenidos, también permite determinar si hubo un cambio y estimar en qué momento aproximadamente ocurrió. Estas capacidades se pueden usar combinándolas, con los gráficos de control CUSUM y EWMA.



Ahora, es necesario realizar un estudio previo de los datos antes de aplicar las herramientas para monitorear el proceso, de modo de detectar los cambios sin proporcionar falsas alarmas y suministrar confiabilidad al estudio. En el análisis previo de los datos existen diferentes metodologías, siendo la de mayor interés el Análisis Exploratorio de Datos propuesto por John Tukey; donde el analista realiza estudios gráficos y pruebas estadísticas para determinar el comportamiento que siguen los datos, además de reconocer las posibles anomalías presentes en la muestra.

La construcción de los gráficos de control antes mencionados, implica efectuar varios análisis, comenzando con el estudio de estacionariedad y luego un estudio de autocorrelación, aplicando pruebas estadísticas para detectar la presencia del mismo.

El objetivo general de esta investigación es el desarrollo de una herramienta computarizada para el análisis de datos provenientes de procesos productivos y su adaptación para la construcción de los gráficos de control EWMA, CUSUM y Análisis de Punto de Cambio, con el fin de proporcionarle al usuario una herramienta automatizada con acceso a una información detallada y depurada de valores que tienden a variar el comportamiento de la muestra, aumentando de esta manera la confiabilidad al estudio que realiza el analista.



El presente trabajo especial de grado consta de cinco capítulos: En el capítulo I se plasman los objetivos, el planteamiento del problema, el alcance y la justificación de la investigación. En el capítulo II se exponen los fundamentos teóricos. En el capítulo III se explica la metodología llevada a cabo durante toda la investigación para lograr cumplir con los objetivos. En el capítulo IV se expone de una forma detallada la secuencia lógica del programa desarrollado y se adapta al lenguaje de programación MATLAB<sup>®</sup>. En el capítulo V a través de una serie de ejemplos teóricos y de la vida real se aplica la herramienta, validando el programa desarrollado y se finaliza con la presentación de las conclusiones y recomendaciones.



## 1. EL PROBLEMA.

### 1.1 PLANTEAMIENTO DEL PROBLEMA.

La estadística según la American Statistical Association (A.S.A) (2012) es la ciencia de la recolección, análisis y presentación de datos. Los estadísticos contribuyen a la investigación científica mediante la aplicación de sus conocimientos para el diseño de encuestas y experimentos, la recopilación, procesamiento y análisis de los datos; interpretando los resultados obtenidos; además, pueden aplicar sus conocimientos en una gran variedad de áreas temáticas, como la biología, la economía, la ingeniería, la medicina, la psicología, la educación y los deportes.

El origen de la estadística se remonta hacia el año 3050 a.c. en el antiguo Egipto cuando los faraones lograron recopilar datos relativos a la población y riqueza del país, con el propósito de preparar la construcción de las pirámides. Luego hacia el año 3000 a.c. cuando los babilonios utilizaban pequeñas tablillas de arcilla para recopilar datos sobre la producción agrícola y los productos obtenidos mediante el trueque. De manera similar ocurrió en Israel, China y Roma durante los periodos siguientes.

Pero no es hasta mediados del siglo XVII cuando Juan Graunt nacido en Londres quien era comerciante, organizó las cifras de mortalidad para estimar





la población de la ciudad en la forma que hoy se conoce como estadística descriptiva, la cual fue publicada como “*Observaciones naturales y políticas hechas sobre la tasa de mortalidad*” (1662).

Con el fin de la segunda guerra mundial se afianzó más la industria y el comercio, siendo necesario transformar la producción para suplir las necesidades de la época, surgiendo contribuciones a los fundamentos estadísticos a la teoría de probabilidades.

En todos los aspectos de la vida, y más aún en el mundo de los negocios, una diversidad asombrosa de datos está disponible para el reconocimiento y la aproximación analítica, de allí que las decisiones hoy en día están soportadas por datos analizados. Actualmente gerentes de negocios y profesionales, son más exigidos a justificar sus decisiones basándose en la información proporcionada por datos y necesitan sistemas de soportes de decisiones basadas en modelos estadísticos.

Para apoyar la toma de decisiones, los datos deben ser precisos y confiables para que los resultados reflejen el comportamiento real del proceso, pero la realidad es que al trabajar con datos, producto de algún proceso de medición o conteo y de la captura de los mismos, generalmente, se producen valores “extraños” como resultado de errores de captura, o de las operaciones aritméticas de quienes operan con ellos. Dentro de estas observaciones se



presentan lo que en estadística se conoce como “outliers”. Según Hawkins (1980), outliers es una observación atípica y/o errónea, que presenta un comportamiento muy diferente con respecto al resto los datos afectando al análisis que se desea realizar sobre las observaciones experimentales. Los outliers aparecen principalmente debido a que las hipótesis de independencia e igualdad en la distribución de las observaciones rara vez son satisfechas en la práctica.

El examen exploratorio de los datos es un paso necesario que lleva tiempo y que habitualmente se descuida por parte de los analistas de datos. Las tareas implícitas en dicho examen pueden parecer insignificantes y sin consecuencias a primera vista, pero son una parte esencial de cualquier análisis estadístico.

Dentro de las herramientas que existen para el estudio de los procesos, se encuentran los gráficos de control, que permiten la detección de cambios en los parámetros de las variables de calidad del producto o servicio. Los más usados son presentados por Shewhart (1931), éstos han mostrado tener gran capacidad para detectar cambios en la media o la varianza del proceso de magnitudes iguales o superiores a dos veces el sigma de la variable en estudio.

Para la detección de cambios menores a dos veces sigma se han desarrollado distintas técnicas que permiten determinar estos pequeños cambios,



de estos estudios nacen los gráficos de control CUSUM (Sumas Acumulativas) propuestos por Page (1954), y los gráficos EWMA (Promedio Móvil Ponderado Exponencialmente) propuesto por Roberts (1959), los cuales a través de simulaciones concluyeron que estos gráficos son más eficientes que los de Shewhart para detectar pequeños cambios en la media del proceso. Esta información permite una mejor toma de decisiones sobre las correcciones necesarias en el proceso; además de lo antes mencionado también se requiere conocer cuáles fueron las causas, el número de veces y en qué instante ocurrió aproximadamente algún cambio en el proceso. Esto es algo que no proporcionan los gráficos de control por si solos. Taylor (2000a) propone el Análisis de Punto de Cambio como una herramienta efectiva para dar respuesta a las interrogantes planteadas.

El Análisis Exploratorio de Datos es una aproximación al análisis de los datos que posterga los supuestos usuales acerca de qué tipo de modelo pertenecen los datos, permitiendo que estos revelen su estructura subyacente y el modelo. No son simples conjuntos de técnicas, es una filosofía en cuanto a la forma en que se realiza la disección de un conjunto de datos, lo que se busca y cómo se interpreta.

Dada la continua búsqueda de mejora en los controles de procesos en las empresas, aunado a la adquisición de nuevos conocimientos por parte de los alumnos de la escuela de Ingeniería Industrial, se integra el Análisis



Exploratorio de Datos con los gráficos de control que detecten cambios menores a dos sigmas para tener una visión de los procesos de manera general. Es por ello que surgió la siguiente interrogante:

¿Cómo diseñar una metodología computarizada que permitan el análisis de la data suministrada y la construcción de un gráfico de control?

## **1.2 OBJETIVOS.**

### **1.2.1 OBJETIVO GENERAL.**

Desarrollar una metodología computarizada para el análisis de datos provenientes de procesos productivos y la construcción de los gráficos de control EWMA, CUSUM y Análisis de Punto de Cambio.

### **1.2.2 OBJETIVOS ESPECÍFICOS.**

1) Enumerar los supuestos estadísticos que se deben cumplir para la construcción de los gráficos de control EWMA, CUSUM y Análisis de Punto de Cambio.



- 2) Diseñar una metodología que permita construir gráficos de control estadístico a partir del análisis del comportamiento de los datos proveniente de procesos productivos.
- 3) Diseñar una herramienta computarizada siguiendo la metodología propuesta.
- 4) Comprobar la efectividad y uso de la herramienta computarizada utilizando datos teóricos y reales que demuestren su aplicabilidad.

### **1.3 ALCANCE DE LA INVESTIGACIÓN.**

El alcance de esta investigación se orienta a la elaboración de una metodología y una herramienta computarizada que permitan integrar el análisis y depuración de datos con la construcción de gráficos de control CUSUM, EWMA y Análisis de Punto de Cambio.

A la data suministrada se le efectuó una serie de pruebas donde se determinó la naturaleza de los datos, identificando comportamientos y realizando transformaciones necesarias para su adaptación a los gráficos de control antes mencionados.

Para la elaboración de la herramienta computarizada se utilizó el lenguaje de programación MATLAB ya que está diseñado especialmente para realizar



cálculos numéricos con vectores y matrices; además de poseer la capacidad de realizar gráficos en dos y tres dimensiones.

#### 1.4 JUSTIFICACIÓN.

El presente trabajo propone una metodología y herramienta computarizada que permite la extracción de información a partir de una fuente de datos. Para ello, se emplearon técnicas estadísticas estándares y gráficos de control para evaluar su comportamiento, ya que es común encontrar datos atípicos que tienen su origen en intervenciones inesperadas y que influyen en la confiabilidad de los resultados, obstaculizando la toma de decisiones.

El análisis de datos proporciona una evaluación preliminar de los mismos mediante métodos que permiten organizarlos, prepararlos e identificar los casos atípicos, detectando fallas en el diseño; con esta data depurada se puede realizar cualquier estudio estadístico aumentando la confiabilidad en los resultados.



El aporte realizado a partir del presente trabajo permite que el estudiante de Ingeniería Industrial tenga a su disposición una herramienta de fácil entendimiento y uso práctico que podrá aplicar para la elaboración de estudios de control de calidad mucho más eficientes que los aprendidos en la asignatura destinada para tal fin, siendo un incentivo para la inclusión de dichos términos y modelos dentro del pensum académico o para estudios posteriores por parte de profesores o estudiantes. Además, permitió a los autores cumplir con el último requisito académico exigido para la obtención del título de Ingeniero Industrial, además de adquirir conocimientos y realizar inferencias con respecto al comportamiento de los datos, para su utilización en los gráficos de control.



## 2. MARCO TEÓRICO.

### 2.1 BASES TEÓRICAS.

#### 2.1.1. NATURALEZA Y FUENTES DE INFORMACIÓN PARA EL ANÁLISIS DE DATOS.

El éxito de cualquier análisis de datos depende de la disponibilidad de la información apropiada. Es muy importante dedicar algún tiempo a estudiar la naturaleza, las fuentes y las limitaciones de los datos que se pueden encontrar en el análisis empírico.

##### 2.1.1.1. TIPOS DE DATOS.

Según Gujarati (2004), puede haber tres tipos de datos disponibles para el análisis empírico: **series de tiempo**, **series de corte transversal** e información **combinada** (combinación de series de tiempo y series de corte transversal).

##### 2.1.1.1.1. SERIES DE TIEMPO.

Según Gujarati (2004), una serie de tiempo es un conjunto de observaciones sobre los valores que toma una variable en diferentes momentos





de tiempo. Tal información debe ser recopilada a intervalos regulares, es decir, en forma diaria, semanal, mensual, trimestral, entre otras. Aunque la información de series de tiempo se utiliza mucho en estudios econométricos, éstas presentan algunos problemas especiales para los econométricos, ya que suponen que éstas son estacionarias. Aunque es pronto para introducir el significado técnico preciso de serie estacionaria, informalmente hablando, *una serie de tiempo es estacionaria cuando la media y su varianza no varían sistemáticamente con el tiempo.*

Según Hernández (2007), una serie de tiempo se puede definir como un conjunto de observaciones de una variable en distintos instantes del tiempo, las cuales son obtenidas a intervalos de tiempo regulares (horas, días, semanas, meses, etc.). Particularmente en los procesos industriales en el área de ingeniería abundan muchos ejemplos, es decir, se consideran series de tiempo muchos datos que provienen de la industria como lo son: la medición de una característica de calidad (viscosidad, temperatura, longitud del producto, conformidad con las especificaciones), el control de las ventas, entre otros.

Los objetivos del análisis de series de tiempo son diversos, en el cual se puede destacar el control de un proceso, la predicción, la simulación de un proceso, etc. Chatfield (2003), plantea el comienzo del estudio con la representación de la data a través de la selección de un modelo probabilístico



hipotético. Luego de escoger el modelo (o la familia de modelos), se pueden estimar los parámetros verificando el ajuste de la data al modelo probabilístico y en muchos casos, se puede usar el modelo ajustado para pronosticar (entendiendo el mecanismo que genera la serie).

#### **2.1.1.1.2. INFORMACIÓN DE CORTE TRANSVERSAL.**

La información de corte transversal consiste en datos de una o más variables recogidos *en el mismo momento del tiempo*, tales como el censo de población realizado por la Oficina del Censo cada 10 años. De igual manera, la información de series de tiempo crea sus propios problemas especiales debido al factor estacionario; la información de corte transversal también tiene sus propios problemas, específicos de heterogeneidad.

#### **2.1.1.1.3. INFORMACIÓN COMBINADA.**

Los datos agrupados tienen elementos de series de tiempo y de corte transversal reunidos.

#### **2.1.1.1.4. DATOS LONGITUDINALES DE PANEL O DE MICROPANEL.**

Hay un tipo especial de datos agrupados, en la cual la misma unidad de corte transversal (una familia o empresa) es encuestada a través del tiempo.



## **2.1.2. ANÁLISIS EXPLORATORIO DE DATOS**

Según Tukey (1977), es el proceso en el que utilizando un conjunto de técnicas estadísticas, se busca evaluar la naturaleza de los datos, su comportamiento en una serie de tiempo y las relaciones existentes entre variables. Su finalidad es conseguir un entendimiento básico de los datos y de las relaciones existentes entre las variables analizadas mediante métodos sistemáticos sencillos para recoger, organizar, diagnosticar y preparar los datos, detectar fallos en el diseño, tratamiento y evaluación de datos atípicos.

### **2.1.2.1. FASES DEL ANÁLISIS EXPLORATORIO DE DATOS.**

#### **2.1.2.1.1. ANÁLISIS ESTADÍSTICO GRÁFICO**

Con el fin de tener una idea inicial de la información contenida en el conjunto de datos así como detectar la existencia de posibles errores en la codificación de los mismos, se procede a realizar un análisis gráfico para observar de manera visual el comportamiento de los datos en el proceso productivo del cual fueron tomados.

#### **2.1.2.1.2. ESTUDIO DE NORMALIDAD**

Gran cantidad de métodos estadísticos se basan en el supuesto de normalidad de la variable en estudio. De hecho, si la falta de normalidad de



la variable es significativa, muchos de los contrastes utilizados en los análisis estadísticos no son válidos.

Existen varios métodos para evaluar la normalidad de un conjunto de datos que pueden dividirse en dos grupos: los métodos gráficos y los contrastes de hipótesis.

#### **2.1.2.1.2.1. MÉTODOS GRÁFICOS.**

El método gráfico univariante más simple para diagnosticar la normalidad es una comprobación visual del histograma que compare los valores de los datos observados con una distribución normal. Aunque atractivo por su simplicidad, este método es problemático para muestras pequeñas, donde la construcción del histograma puede distorsionar la representación visual de tal forma que el análisis sea poco fiable.

##### **2.1.2.1.2.1.1. ELABORACIÓN DEL HISTOGRAMA.**

Walpole (1999), describe en su obra los pasos a seguir para la construcción de histogramas de la siguiente manera:

- Obtener una muestra y los valores de la variable que se estudia. Mínimo 30 datos. Es recomendable utilizar una hoja de registros.



- Calcular el rango o amplitud de los datos (diferencia entre el mayor y el menor de los datos).
- Determinar el ancho de cada intervalo que servirá para construir el histograma.

Es recomendable que el histograma tenga de 5 a 20 barras. Una buena aproximación del número de intervalos aconsejable se obtiene calculando la raíz cuadrada del número de datos

Se aconseja que el tamaño o amplitud de intervalo tenga un grado de aproximación no mayor a aquel con el que se registran los datos. Se realiza de la siguiente manera:

- Establecer los límites o fronteras de cada clase, es decir, los valores de inicio y terminación de cada intervalo.
- Construir la tabla. La tabla se puede construir de diferentes formas pero hay que tener en cuenta que el primer intervalo debe contener el menor de los datos y el último el mayor. Asimismo, la presentación de los datos en la tabla no debe generar confusiones acerca del intervalo que contiene cada dato. En lo posible, todos los intervalos deben tener el mismo ancho. Es usual que en la primera columna se registre el número de orden de cada clase, en la segunda se escriban los



intervalos, en la tercera las marcas de clase y en la cuarta las frecuencias.

- Graficar el histograma. En lo posible dar una presentación tal que la altura sea aproximadamente del ancho de la gráfica.

El histograma de frecuencias en sí es una sucesión de rectángulos construidos sobre un sistema de coordenadas cartesianas de la manera siguiente:

- Las bases de los rectángulos se localizan en el eje horizontal. La longitud de la base es igual al ancho del intervalo.
- Las alturas de los rectángulos se registran sobre el eje vertical y corresponden a las frecuencias de las clases.
- Las áreas de los rectángulos son proporcionales a las frecuencias de las clases.
- Los histogramas pueden estar referidos a las frecuencias absolutas, a las frecuencias acumuladas o a las frecuencias relativas o porcentuales.



Otra técnica descriptiva para comprobar la normalidad es la gráfica de probabilidad normal. En una gráfica de probabilidad normal, las observaciones de un conjunto de datos se ordenan y luego se grafican contra los valores esperados estandarizados de las observaciones bajo el supuesto de que los datos están distribuidos normalmente. Si los datos en verdad tienen una distribución normal, una observación será aproximadamente igual a su valor esperado. Por tanto, una tendencia lineal (de línea recta) en la gráfica de probabilidad normal sugiere que los datos provienen de una distribución aproximadamente normal, en tanto que una tendencia no lineal indica que los datos no son normales.

#### 2.1.2.1.2.1.2. ELABORACIÓN DE GRÁFICO DE PROBABILIDAD NORMAL.

Según Walpole (1999), la construcción del gráfico de probabilidad normal para un conjunto de datos es de la siguiente manera:

- Se ordenan los  $n$  datos de forma ascendente.
- Se asigna una posición a cada dato, siendo la posición  $1$  para el dato más pequeño y la posición  $n$  para el dato más grande. Esto hace que los datos ordenados se denoten por  $X_1, \dots, X_n$ .
- Se calcula la posición de representación para el dato que ocupa el lugar  $i$  en los datos ordenados, asignándole una frecuencia acumulada  $F_i$  dada por la siguiente expresión:



$$F_i = 100 \frac{3i - 3/8}{n + 1/4} \quad \text{con } i = 1, \dots, n \quad (2.1)$$

- Se representan sobre el gráfico los puntos.
- Se ajusta una línea recta a los datos minimizando las desviaciones entre recta y puntos representados.

Sobre el gráfico resultante se podrá admitir que los datos siguen una distribución normal cuando la nube de puntos obtenida en el gráfico se ajuste bien a la recta. En otro caso habrá que rechazar la normalidad de los datos.

#### 2.1.2.1.2.2. CONTRASTES DE HIPÓTESIS

La segunda de las formas para comprobar la normalidad de una distribución se efectúa a través de un contraste de hipótesis. Mohd y Bee (2011) concluyen que el contraste de Shapiro y Wilks es, en términos generales, el más conveniente en muestras pequeñas, mientras que el contraste de Kolmogorov-Smirnov, es adecuado para muestras grandes. Además de las pruebas mencionadas anteriormente la prueba de Lilliefors es un caso especial del contraste Kolmogorov-Smirnov.

#### 2.1.2.1.3. IDENTIFICACIÓN DE VALORES ATÍPICOS (OUTLIERS)

##### 2.1.2.1.3.1. TIPOS DE OUTLIERS





Según Beckman y Cook (1983) los outliers pueden ser de distintos tipos:

- Observación discordante es aquella observación que parece sorprendente o discrepante para el investigador.
- Observación contaminante es aquella que no proviene de la población en estudio.
- Observación influyente es aquella que al ser excluida del análisis de datos, altera sustancialmente rasgos importantes de dicho análisis.

#### **2.1.2.1.3.2. IDENTIFICACIÓN DE OUTLIERS**

Barnett y Lewis (1978) considera el estudio de los datos univariados y multivariados, modelos lineales, regresión y de series de tiempo, en particular la influencia de los valores atípicos en los datos estadísticos.

Marks y Rao (1979) realizaron un procedimiento de estimación de los datos que contienen los valores extremos con un cambio unidireccional en la media. Los datos obtenidos en el experimento a menudo incluyen



observaciones espurias, que deben ser tomadas en cuenta en el análisis de datos.

Beckman y Cook (1983) revisan las técnicas de detección para múltiples outliers. El trabajo sobre outliers en datos circulares, el análisis de discriminante, el diseño experimental, datos multivariantes, modelos generalizados lineales, otras distribuciones además de la normal, y la serie de tiempo también es revisada.

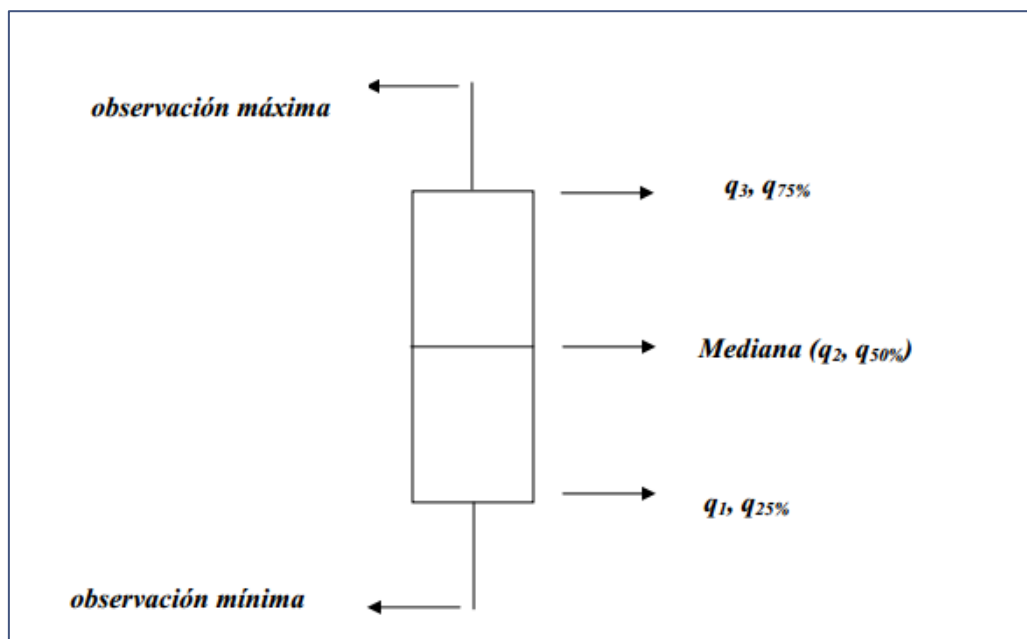
Para la detección de los datos atípicos se pueden emplear múltiples herramientas como los gráficos de caja, los diagramas de control y pruebas estadísticas.

#### **2.1.2.1.3.2.1. ELABORACIÓN DE DIAGRAMA DE CAJA.**

Los diagramas de caja o boxplots son un aporte original de Tukey (1977) como herramienta de la estadística descriptiva para visualizar rápidamente las características de tendencia central, dispersión, localización y simetría de un conjunto de datos, o para comparar estas características a través de varios subgrupos de una población.



Cuando se estudia una sola variable (caso univariante), el diagrama de cajas consta de los siguientes elementos:



**Figura 2.1** Representación gráfica del Diagrama de Caja.

**Fuente:** Elaboración propia.

#### 2.1.2.1.3.2.2. PRUEBA DE ROSNER



Según Hawkins (1980), la prueba de Rosner (1975) para detectar  $k$  valores atípicos puede ser utilizada cuando el número de los datos de referencias son 25 o más. Los datos se ordenan de forma ascendente y se determina la media y la desviación estándar. El procedimiento exige remover la observación  $x$ , siendo este el más alejado de la media. Entonces una prueba estadística  $R$ , vendrá de la siguiente manera:

La hipótesis nula será:

*H<sub>0</sub>: No hay presencia de outliers en la muestra.*

*H<sub>1</sub>: Hay  $r$  outliers en la muestra.*

$$R_{i+1} = \frac{\max |X_i - \bar{X}_i|}{S_i}$$

(2.2)

Donde  $\bar{X}$  y  $S$  denotan la media muestral y la desviación estándar muestral, respectivamente.

Se remueve la observación que maximiza  $|X_i - \bar{X}_i|$  luego se vuelve a calcular la estadística anterior con  $n - 1$  observaciones. Se repita este proceso hasta las  $r$  observaciones que se han eliminado. Esto resulta en la prueba estadística  $r$   $R_1, R_2, \dots, R_r$ .



El nivel de significancia será  $\alpha$  y en correspondencia con la prueba estadística  $r$ , se establece los valores  $r$  críticos de la siguiente forma:

$$\lambda_i = \frac{(n-i)t_{p,n-i-1}}{\sqrt{(n-i-1+t_{p,n-i-1}^2)(n-i+1)}}$$

(2.3)

donde  $i = 1, 2, \dots, K, r$ ,  $t_{p,v}$  es el punto de porcentaje de la **distribución  $t$**  con  $v$  grados de libertad y

$$p = 1 - \frac{\alpha}{2(n-i+1)}$$

(2.4)

Si  $R$  es menor que el valor crítico, la hipótesis nula no puede ser rechazada, y por lo tanto no hay valores atípicos. Si  $R$  es mayor que el valor crítico, se rechaza la hipótesis nula y la presencia de los  $k$  valores atípicos se acepta.

El número de valores atípicos se determina encontrando la mayor  $i$  tal que

$$R_i > \lambda_i$$

### 2.1.3. PROCESOS ESTOCÁSTICOS



Pérez (2007), define el proceso estocástico como una familia de variables aleatorias  $\{Y_t\}$ , donde  $t$  denota el tiempo, tales que para cualquier elección finita de valores de  $t_1, t_2, t_3, \dots, t_n$  existe la distribución de probabilidad conjunta correspondiente a las variables aleatorias  $Y_{t_1}, Y_{t_2}, \dots, Y_{t_n}$ . También clasifica los procesos estocásticos como estacionarios y no estacionarios.

Según Pérez (2007), una de las formas de describir un proceso estocástico es con los momentos poblacionales, particularmente con los de primer y segundo orden.

$$\text{i. } E(Y_t) = \mu_t, \quad " t$$

(2.5)

$$\text{ii. } V(Y_t) = \sigma_t^2, \quad " t$$

(2.6)

$$\text{iii. } \gamma_{t,s} = Cov(Y_t, Y_s) = E[(Y_t - \mu_t)(Y_s - \mu_s)] \quad " t, s$$

(2.7)

Es decir, se especifica como la media, la varianza y la covarianza dependen del tiempo.

### 2.1.3.1. PROCESOS ESTOCÁSTICOS ESTACIONARIOS.



Según Pérez (2007), un tipo de proceso estocástico es el llamado proceso estocástico estacionario. En términos generales, se dice que un proceso estocástico es estacionario si su media y su varianza son contantes en el tiempo y si el valor de la covarianza entre dos periodos depende solamente de la distancia o rezago entre estos dos periodos de tiempo y no del tiempo en el cual se ha calculado la covarianza.

### 2.1.3.2. PROCESOS ESTOCÁSTICOS NO ESTACIONARIOS

Aunque el interés se centra en las series de tiempo estacionarias, a menudo se encuentran series de tiempo no estacionarias, siendo el clásico ejemplo el modelo de caminata aleatoria: **sin variaciones** (con termino constante), y **con variaciones** (es decir, sin termino constante o de intersección).

### 2.1.3.3. CAMINATA ALEATORIA SIN VARIACIONES.

Supóngase que  $u_t$  es un término de error con ruido blanco, el cual se caracteriza por tener: media nula, varianza constante y autocovarianzas nulas en todos los retardos distintos de cero. Se dice entonces que la serie  $Y_t$  es caminata aleatoria si:



$$Y_t = Y_{t-1} + u_t$$

(2.8)

En el modelo de caminata aleatoria, el valor  $Y$  en el tiempo  $t$  es igual a su valor en el tiempo  $(t-1)$  más un factor aleatorio; por tanto, es un modelo AR.

#### 2.1.3.4. CAMINATA ALEATORIA CON VARIACIONES.

Se modificara la ecuación (2.8) de la siguiente forma:

$$Y_t = \delta + Y_{t-1} + u_t$$

(2.9)

Donde  $\delta$  se conoce como el parámetro de variación. El nombre de variación proviene del hecho que  $Y_t$ , varia hacia arriba o hacia abajo, dependiendo si  $\delta$  es positiva o negativa.

#### 2.1.3.5. PROCESO ESTOCÁSTICO DE RAÍZ UNITARIA.

Se escribirá el modelo de caminata aleatoria (2.8) como:





$$Y_t = \rho Y_{t-1} + u_t \quad - 1 \leq \rho \leq 1$$

(2.10)

Si  $\rho = 1$ , la ecuación (2.10) se convierte en un modelo de caminata aleatoria sin variaciones y se tiene lo que se conoce como problema de raíz unitaria; es decir, se enfrenta a una situación de no estacionariedad.

Sin embargo, si  $\rho$  es menor que 1, entonces se puede demostrar que la serie de tiempo  $Y_t$  es estacionaria.

#### 2.1.4. PROCESOS ESTACIONARIOS.

Chatfield (2003), define el proceso estacionario como aquellos procesos en los cuales no existe un cambio significativo en la media (no existe tendencia), no hay cambios significativos en la varianza y los componentes estacionales de la serie de tiempo han sido eliminados rigurosamente. Cuando un proceso es estacionario, se puede decir que está bajo equilibrio estadístico (comúnmente llamado así), y esto no es más que dividir el tiempo en intervalos de forma tal que las propiedades estadísticas no cambien en el tiempo.

Es importante resaltar el hecho de que la mayoría de los modelos probabilísticos que analizan series de tiempo, se enfocan en las estacionarias,



es por esto que en los casos donde no se cuenta con una, la serie de tiempo no estacionaria se debe convertir en estacionaria. Por esa razón, es mucho interés para el análisis de series de tiempo el concepto de estacionariedad. Al momento de caracterizar completamente un proceso estocástico, es necesario conocer la función de densidad conjunta de todas las variables aleatorias involucradas. Sin embargo, en la práctica no es muy realista pensar que esto se pueda lograr con una serie de tiempo.

Supóngase que la media de  $Y_t$  es  $\mu_t$ , ya que se asume que la media del error aleatorio,  $a_t$  es cero, es decir, el  $E(a_t) = 0$ , para toda  $t$ . Por consiguiente, si la media del proceso no depende del tiempo, implica que, aun cuando durante un cierto lapso de tiempo el proceso se aleje de la media, el proceso se acercará a ella eventualmente.

Con el fin de limitar y entender mejor los conceptos de estacionariedad, se estudian los procesos fuertemente estacionarios y débilmente estacionarios. Un proceso fuertemente estacionario es aquel cuya distribución conjunta de  $X(t_1), K, X(t_n)$ , es la misma distribución conjunta de  $X(t_1 + \tau), K, X(t_n + \tau)$ , para toda observación  $\{(t_1, K, t_n), \tau\}$ . Lo que lleva a pensar que la serie va a mostrar el mismo comportamiento, sin importar el momento en que se observe el proceso. Por lo tanto, para estos procesos, el nivel de la serie y su variabilidad no dependen del tiempo, y con respecto a la covarianza, no existe dependencia en el tiempo.



Un proceso débilmente estacionario o de segundo orden, es aquel donde su media es constante y su función de autocovarianza depende de la separación ( $k$ ) que hay entre variables (esta separación se conoce con el nombre de Lag). Esta definición implica que tanto la varianza como la media son constantes y su valor debe ser finito.

#### **2.1.5. ESTACIONARIEDAD.**

Uno de los propósitos del control de calidad es inducir un proceso estable, pero en muchas ocasiones se asume estacionariedad a priori; Box, Jenkins y Reinsel (1994), afirma que en la práctica la mayoría de series de tiempo con las que se trabaja no son estacionarias, sin embargo, no debe representar un problema, ya que existen diversas formas para solucionar dicha complicación y convertir a las series de tiempo no estacionarias, en estacionarias.

##### **2.1.5.1. PRUEBAS DE ESTACIONARIEDAD.**

A pesar de que existen muchas pruebas de estacionariedad, sólo se presentan dos formas de hacerlo: a través de la prueba de raíz unitaria de **Dickey-Fuller** y de la Prueba de raíz unitaria de **Phillips-Perron**.



### 2.1.5.1.1. PRUEBA DE RAÍZ UNITARIA DE DICKEY-FULLER (DF).

Gujarati (2004), presenta el test de raíces unitarias propuesto por Dickey-Fuller (1979) partiendo del modelo:

$$Y_t = \rho Y_{t-1} + u_t \quad (2.11)$$

Donde:

$u_t$  = Es el término de error con ruido blanco.

$\rho$  = Es un coeficiente de autocorrelación, por lo tanto su valor debe estar entre:  $-1 \leq \rho \leq 1$

Si  $\rho = 1$  se presenta el caso de raíz unitaria. Ahora, la ecuación (2.11) puede ser expresada alternativamente restando  $Y_{t-1}$  en ambos lados de la ecuación (2.10) para obtener la siguiente regresión:

$$\begin{aligned} Y_t - Y_{t-1} &= \rho Y_{t-1} - Y_{t-1} + u_t \\ Y_t - Y_{t-1} &= (\rho - 1)Y_{t-1} + u_t \end{aligned} \quad (2.12)$$

Que puede ser escrito alternativamente como:

$$\Delta Y_t = \delta Y_{t-1} + u_t \quad (2.13)$$

Donde:



$\delta = (\rho - 1)$  y  $\Delta$ , es el operador de la primera diferencia.

Es decir, en lugar de usar la ecuación (2.11), se calcula la ecuación (2.13) y se prueba la hipótesis nula de que  $\delta = 0$ , si  $\delta = 0$  quiere decir que  $\rho = 1$ ; es decir, se tiene una raíz unitaria, por lo tanto, significa que la serie de tiempo estudiada no es estacionaria. De lo contrario, si el estimador de  $\delta$  es negativo y significativamente diferente de cero, se rechaza la hipótesis nula.

El procedimiento de implantar la prueba de DF, involucra diversas decisiones. Porque se puede observar en los procesos de caminatas aleatorias pueden tener variaciones o no, o posiblemente pueden tener tendencias determinísticas y estocásticas. Por lo tanto, la prueba se estima bajo tres diferentes hipótesis nulas como sigue:

$Y_t$  es una caminata aleatoria:

$$DY_t = \delta Y_{t-1} + u_t \quad (2.14)$$

$Y_t$  es una caminata aleatoria con variaciones:

$$DY_t = \beta_1 + \delta Y_{t-1} + u_t \quad (2.15)$$

$Y_t$  es una caminata aleatoria con variaciones alrededor de una tendencia estocástica:



$$DY_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + u_t \quad (2.16)$$

Donde  $t$  es el tiempo o tendencia variable. En cada caso, la hipótesis nula es que  $\delta = 0$ , el cual quiere decir que es una raíz unitaria como se explicó anteriormente. Si la hipótesis nula es rechazada, significa que  $Y_t$  es una serie de tiempo estacionaria con media cero en el caso de la ecuación (2.13), que  $Y_t$  es estacionaria con una media distinta de cero ( $= B_1/(1-\rho)$ ) en el caso de la ecuación (2.14), y que  $Y_t$  es estacionaria alrededor de una tendencia determinística en el caso de la ecuación (2.16).

#### 2.1.5.1.2. PRUEBA DE RAÍZ UNITARIA DE PHILLIPS-PERRON (PP).

Según Gujarati (2004), el contraste de Phillips y Perron (1988) es un método estadístico no paramétrico para evitar la correlación serial de orden elevado en los términos de error, de modo tal, que no se le añadan términos de diferencia rezagados.

En ésta prueba se aumenta a las ecuaciones de DF los valores rezagados de la variable dependiente  $DY_t$ , debido a que se supone que el término de error se encuentra correlacionado. Por tanto, la prueba consiste en estimar la siguiente regresión:

$$DY_t = \beta_1 + \beta_2 t + \delta Y_{t-1} + \alpha_i \hat{\alpha}_{i=1}^m DY_{t-1} + \varepsilon_t \quad (2.17)$$



Donde:

$\varepsilon_t$  : Es el término de error puro con ruido blanco. En esta prueba se sigue la misma distribución asintótica que el estadístico DF.

### **2.1.5.2. TRANSFORMACIÓN PARA OBTENER SERIES DE TIEMPO ESTACIONARIAS.**

Carrasquel, González y Rodríguez (2001), sugieren en su libro que en los casos donde la serie de tiempo no sea estacionaria, se debe transformar inicialmente la serie de tiempo original de modo tal que se consiga la estacionariedad en la varianza y seguidamente en la media.

#### **2.1.5.2.1 CORRECCIÓN DE LA NO ESTACIONARIEDAD EN LA VARIANZA.**

Box-Cox (1964), hace el planteamiento de que existe una familia de transformaciones que generalmente consigue la transformación adecuada de una serie de tiempo para hacerla estacionaria en la varianza, la familia del cual se hace referencia es la siguiente:

$$Y_{(\lambda)} = \frac{Y^\lambda - 1}{\lambda} \quad (2.18)$$



La cual depende del parámetro  $\lambda$ , se prueba  $\lambda$  con distintos valores (por lo general  $\lambda = 1; 0.5; 0; -0.5; -1$ , aunque pueden ser otros) y el valor de  $\lambda$  que produzca un gráfico de una nube de puntos aproximadamente paralela al eje de las abscisas será el tomado en consideración. Con el fin de aclarar un poco más el concepto, a continuación se ilustrará lo antes mencionado:

**Tabla 2.1.** Distintos valores de  $\lambda$  usados en la transformación de Box-Cox.

Valor de $\lambda$	Transformación	Serie Transformada	Resultado
$\lambda = 1$	Sin transformación	$Y_{(1)} = \frac{Y^1 - 1}{1}$	$Y$
$\lambda = 0.5$	Raíz cuadrada	$Y_{(0.5)} = \frac{Y^{0.5} - 1}{0.5}$	$\sqrt{Y}$
$\lambda = 0$	Logarítmica	$Y_{(0)} = \frac{Y^0 - 1}{0}$	$\ln(Y)$
$\lambda = -0.5$	Inversa de la raíz cuadrada	$Y_{(-0.5)} = \frac{Y^{-0.5} - 1}{-0.5}$	$\frac{1}{\sqrt{Y}}$
$\lambda = -1$	Inversa	$Y_{(-1)} = \frac{Y^{-1} - 1}{-1}$	$\frac{1}{Y}$

**Fuente:** Carrasquel, González y Rodríguez (2001).

#### 2.1.5.2.2. CORRECCIÓN DE LA NO ESTACIONARIEDAD EN LA MEDIA.

Luego de asegurarnos que la serie de tiempo obtenida es estacionaria en la varianza, el siguiente paso es lograr la estacionariedad en la media en caso de que no lo sea.





Según Gujarati (2004), existen pruebas estadísticas para verificar la estacionariedad en la media. Si el resultado de la prueba arroja que no es estacionaria en la media, se puede transformar la serie de tiempo realizando unas operaciones matemáticas con el fin de convertirlas en estacionarias; estas operaciones se conocen como diferenciaciones.

De ser necesario se pueden realizar varias diferenciaciones para lograr la estacionariedad en la media. La primera diferenciación se plantea como sigue:

$$DY_t = Y_t - Y_{t-1} \quad (2.19)$$

Donde:

$Y_t$  = observaciones en el tiempo  $t$

$Y_{t-1}$  = observaciones en el tiempo  $t - 1$

Luego de realizar la primera diferenciación se comprueba la estacionariedad, y de no ser estacionaria se aplica la segunda diferenciación de la siguiente forma:

$$DY_t = (Y_t - Y_{t-1}) - (Y_{t-1} - Y_{t-2}) \quad (2.20)$$

Entonces se encuentran entre dos escenarios: El primero corresponde al caso donde la serie de tiempo debe ser diferenciada solamente una vez para que pueda ser estacionaria, aquí se dice que la serie original es integrada de



orden 1, y se denota por  $I(1)$ . En forma similar si la serie original debe ser diferenciada dos veces, es decir, al resultado de la primera diferencia se le aplica la segunda diferencia para poder hacerla estacionaria, se dice que la serie original es integrada de orden 2, o  $I(2)$ .

### **2.1.6. MÉTODO DE BOX & JENKINS.**

Según Hanke y Reitsch (2000), la metodología Box – Jenkins, utiliza un enfoque iterativo de identificación de un modelo útil a partir de un modelo general.

#### **2.1.6.1. METODOLOGÍA DE BOX & JENKINS.**

La aplicación de esta metodología exige las siguientes fases:

- Identificación del proceso subyacente.
- Estimación de los parámetros de los procesos identificados.
- Validación del proceso estimado.

##### **2.1.6.1.1. IDENTIFICACIÓN.**

Una vez logradas las transformaciones necesarias en las series para hacerlas estacionarias en media y varianza, se procede a obtener las



funciones de autocorrelación y de autocorrelación parcial de las series transformadas.

Se identifican los procesos partiendo de las propiedades teóricas de los procesos AR o MA, según los cuales, un proceso  $AR(p)$  tiene una función de autocorrelación parcial que se anula a partir del retardo  $p$ , y un proceso  $MA(q)$  tiene una función de autocorrelación que se anula a partir del retardo  $q$ .

El cuadro adjunto permite resumir las pautas de comportamiento más frecuentes de la función de autocorrelación y la función de autocorrelación parcial en los modelos  $AR(p)$ ,  $MA(q)$  y  $ARMA(p, q)$ .

**Tabla 2.** Regla de decisión para elección de modelos de series de tiempo.

	<b>Función de autocorrelación (FAC)</b>	<b>Función de autocorrelación Parcial (FACP)</b>
<b>AR(p)</b>	Decrecimiento rápido de tipo exponencial y / o sinusoidal.	Se anula para retardos superiores a $p$ .
<b>MA(q)</b>	Se anula para retardos superiores a $q$ .	Decrecimiento rápido de tipo exponencial y/o sinusoidal.
<b>ARMA (p, q)</b>	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones	Los primeros valores iniciales no tienen patrón fijo y van seguidos de una mezcla de oscilaciones sinusoidales



	sinusoidales amortiguadas a partir de $q$ .	amortiguadas a partir de $p$ .
--	---	--------------------------------

**Fuente:** Análisis Econométrico en Eviews.

### 2.1.6.1.2. ESTIMACIÓN

Una vez seleccionados los diferentes procesos posibles que puedan haber generado la serie se procede a la estimación.

### 2.1.6.1.3. VALIDACIÓN

Los modelos estimados deben pasar las pruebas clásicas de estimación de cualquier modelo. En particular:

- Todos sus parámetros deben ser significativos.
- Los residuos deben ser ruido blanco, lo que se puede observar si al hacer el correlograma, sus valores caen dentro de los límites, o mejor, si la prueba asociada al estadístico  $Q$  de Ljung-Box que los acompaña es mayor que  $5\%$ , para realizar los contrastes al  $95\%$  de confianza.

### 2.1.6.2. MODELOS DE SERIES DE TIEMPO MA( $q$ ), AR( $p$ ), ARMA( $p,q$ ).

Existen diversos modelos que son usados para ajustar una serie de tiempo, estos facilitan el estudio del comportamiento de las series porque con



los patrones se estiman sus parámetros y se puede realizar cualquier estudio deseado (pronósticos, estudios de autocorrelación, entre otras). Estos modelos son: MA (q), AR (p) Y ARMA (q, p). En primer lugar se estudiara el modelo de promedio móvil de la siguiente forma:

#### 2.1.6.2.1. MODELO PROMEDIO MÓVIL O MOVING AVERAGE (MA (q)).

Suponga que  $\{Z_t\}$  es un proceso completamente aleatorio con media cero y varianza  $\sigma_Z^2$ . Entonces, un proceso  $\{X_t\}$ , según Chatfield (2003), se dice que es un proceso promedio móvil de orden q MA(q) si:

$$X_t = \beta_0 Z_t + \beta_1 Z_{t-1} + \dots + \beta_q Z_{t-q} \quad (2.21)$$

Donde  $\{\beta_i\}$  son constantes. Las  $Z_t$  son usualmente escaladas para hacer  $\beta_0 = 1$ . Se puede encontrar que la media y varianza son:

$$E(X_t) = 0 \quad (2.22)$$

y



$$\text{Var}(X_t) = \sigma_Z^2 \sum_{i=0}^q \beta_i^2 \quad (2.23)$$

Dado que las  $Z_t$  son independientes. Se tiene también:

$$\gamma(k) = \text{Cov}(X_t, X_{t+k}) \quad (2.24)$$

$$\gamma(k) = \text{Cov}(\beta_0 Z_t + \dots + \beta_q Z_{t-q}, \beta_0 Z_{t+k} + \dots + \beta_q Z_{t+k-q})$$

$$(2.25) \quad \gamma(k) = \begin{cases} 0 & k > q \\ \sigma_Z^2 \sum_{i=0}^{q-k} \beta_i \beta_{i+k} & k = 0, 1, \dots, q \\ \gamma(-k) & k < 0 \end{cases}$$

$$(2.26)$$

Como  $\gamma(k)$  no depende de  $t$ , y la media es constante, el proceso es estacionario de segundo orden para todo valor de  $\{\beta_i\}$ . Adicionalmente, si los  $Z_t$  están normalmente distribuidos, entonces también lo están las  $X_t$  y se tiene un proceso normal estrictamente estacionario.

#### 2.1.6.2.2. MODELOS AUTORREGRESIVOS AR(p).

Según Chatfield (2003), Los procesos autorregresivos se aplican a muchas situaciones, en las que es razonable asumir que el valor presente de la serie de tiempo depende del valor inmediato anterior junto con un error aleatorio. Los procesos autorregresivos se representan bajo la siguiente forma:



Suponga que  $\{Z_t\}$  es un proceso completamente aleatorio con media cero y varianza  $\sigma_Z^2$ . Entonces, un proceso  $\{X_t\}$  es dicho que es un proceso autorregresivo de orden  $p$  si:

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + Z_t \quad (2.27)$$

En caso de que en el proceso  $Z_t$  tenga una media constante dada por  $E(Z_t) = \mu$  para toda  $t$ ; se obtiene la siguiente ecuación:

$$\phi(B)Z_t = a_t, \quad \text{con } Z_t = Z_t - \mu \quad (2.28)$$

El término autorregresivo (AR) que se le da al proceso representado se refiere al hecho de que también puede expresarse así:

$$Z_t = (1 - \phi_1 - \phi_2 - \dots - \phi_p)\mu + \phi_1 Z_{t-1} - \phi_2 Z_{t-2} - \dots - \phi_p Z_{t-p} + a_t \quad (2.29)$$

Si se observa con detalle la estructura de la ecuación representada tiene la forma de una ecuación de regresión lineal, con la característica de que el valor de la variable dependiente  $Z$  en el periodo  $t$ , no depende de los valores de un cierto conjunto de variables independientes, como pasa en el modelo de regresión, sino de sus propios valores observados en periodos anteriores a  $t$ , y ponderados de acuerdo con los coeficientes autorregresivos  $\phi_1, \phi_2, \dots, \phi_p$ .



Es importante entender si el proceso con una ecuación en diferencia alcanzará eventualmente su punto de equilibrio. Tratándose de ecuaciones en diferencia en la que está presente algún elemento aleatorio, no se puede hablar precisamente de convergencia debido a las fluctuaciones aleatorias, aun cuando estas ocurran alrededor del punto de equilibrio. Por lo tanto el concepto de equilibrio estocástico, mejor conocido como estacionalidad, tiene mayor utilidad. Es decir, mientras que en un proceso determinístico se habla de equilibrio, cuando se tiene un proceso estocástico se habla de estacionalidad; de esta manera, un proceso AR podrá ser estacionario o no-estacionario, dependiendo de los valores que tomen las raíces de la ecuación característica  $(\phi(x) = 0)$ , observándose como se desarrolla el proceso autorregresivo. De tal manera que el proceso AR definido por  $\phi(B)$  será estacionario siempre y cuando

$$|g_i| < 1 \quad \text{para } i = 1, 2, \dots, p \quad (2.30)$$

o dicho de otra manera, si y solo si la raíces se encuentran fuera del círculo unitario (plano complejo).

#### 2.1.6.2.2.1. MODELOS AUTORREGRESIVOS DE PRIMER ORDEN AR (1).





Para simplicidad, se examina el caso de primer orden donde  $p = 1$ .  
Entonces,

$$X_t = \alpha_1 X_{t-1} + \dots + Z_t \quad (2.31)$$

El proceso AR(1) es muchas veces llamado el proceso Markov, debido a Russian A. A. Markov. Por sucesiva sustitución la ecuación (2.31) se puede escribir:

$$X_t = \alpha_1 [\alpha_1 X_{t-2} + \dots + Z_{t-1}] + Z_t \quad (2.32)$$

$$X_t = \alpha_1^2 [\alpha_1 X_{t-3} + \dots + Z_{t-2}] + \alpha_1 Z_{t-1} + Z_t \quad (2.33)$$

Y eventualmente se encuentra que  $X_t$  puede ser expresada como un proceso MA de orden infinito en la forma (previando que  $-1 < \alpha < 1$ ):

$$X_t = Z_t + \alpha Z_{t-1} + \alpha^2 Z_{t-2} + \dots \quad (2.34)$$

### 2.1.6.2.3. MODELOS MEZCLADOS (COMBINACIÓN DE MA CON AR, ARMA (p, q)).



Según Chatfield (2003), una clase útil de modelos de series de tiempo es la formada como combinación de los procesos MA y AR. Un proceso promedio móvil autorregresivo contiene  $p$  términos AR y  $q$  términos MA (abreviado sería proceso ARMA  $(p, q)$ ), y es de orden  $(p, q)$ . Está dado por:

$$X_t = \alpha_1 X_{t-1} + \dots + \alpha_p X_{t-p} + Z_t + \beta_1 Z_{t-1} + \dots + \beta_q Z_{t-q} \quad (2.35)$$

#### 2.1.6.2.3.1. ESTIMACIÓN DE PARÁMETROS PARA EL MODELO ARMA $(p, q)$ .

Suponga que un modelo promedio móvil autorregresivo es pensado de forma apropiada para una serie de tiempo dada. Según C. Chatfield (1980), los problemas de estimación para un modelo ARMA son similares a los de un modelo MA y el proceso iterativo es usado. La suma residual de cuadrados puede ser calculada a cada punto sobre una reja satisfactoria de valores de parámetros, y los valores que devuelve la suma de mínimos cuadrados pueden entonces ser evaluados. Alternativamente, algunos tipos de optimización pueden ser usados.

Como un ejemplo, considere el proceso ARMA  $(1,1)$  cuya función de autocorrelación decrece exponencialmente hasta uno. Este modelo puede ser reconocido como apropiado si la muestra de función autocorrelacionada tiene una forma similar. El modelo es dado por:



$$X_{t-\mu} = \alpha_1 (X_{t-1} - \mu) + Z_t + \beta_1 Z_{t-1} \quad (2.36)$$

Dado  $N$  observaciones  $X_1, X_2, \dots, X_N$  se suponen valores para  $\mu, \alpha_1, \beta_1$ , conjunto  $Z_0 = 0$  y  $X_0 = \mu$ , y entonces se calculan los residuales repetitivamente por:

$$Z_1 = X_1 - \mu \quad (2.37)$$

$$Z_2 = x_2 - \mu - \alpha_1(x_1 - \mu) - \beta_1 Z_1 \quad (2.38)$$

$$Z_N = (x_N - \mu) - \alpha_1(x_{N-1} - \mu) - \beta_1 Z_{N-1} \quad (2.39)$$

La suma residual de cuadrados  $\sum_{t=1}^N Z_t^2$  puede entonces ser calculada.

Entonces otros valores de  $\mu, \alpha_1, \beta_1$ , pueden ser probados hasta que la suma mínima residual de cuadrados es encontrado.

### 2.1.7. AUTOCORRELACIÓN.

Kendall y Buckland (1971), definen en su diccionario el término de autocorrelación como la correlación que existe entre los miembros de las series de observaciones ordenadas en el tiempo (es decir, los datos de series de tiempo) o el espacio (como en los datos de corte transversal).



Chatfield (2003), señala que los coeficientes de autocorrelación son una guía muy importante para las propiedades de una serie de tiempo, la cual mide la correlación entre observaciones separadas a diferentes distancias. Frecuentemente, éste coeficiente suministra suspicacia dentro del modelo de probabilidad generada de la data. En el presente trabajo se asume que el lector se encuentra familiarizado con el coeficiente de correlación ordinario, el cual puede ser expresado de la siguiente forma:

Dado  $N$  pares de observaciones de dos variables  $X$  y  $Y$ , el coeficiente de correlación será dado por:

$$r = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}} \quad (2.40)$$

En una serie de tiempo se puede aplicar el mismo concepto de una forma similar, con el fin de ver si las observaciones sucesivas se encuentran correlacionadas. Es decir, como se presenta a continuación:

Dada  $N$  observaciones  $x_1, \dots, x_N$ , en una serie de tiempo discreta se forman  $(N - 1)$  pares de observaciones expresadas como:  $(x_1, x_2), (x_2, x_3), \dots, (x_{N-1}, x_N)$ .



En el cual, respecto a la primera observación de cada par de una variable, y la segunda observación de una segunda variable, el coeficiente de correlación entre  $x_t$  y  $x_{t-1}$  se puede expresar como:

$$r = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})(x_{t+1} - \bar{x}_{(2)})}{\sqrt{\sum_{t=1}^{N-1} (x_t - \bar{x}_{(1)})^2 \sum_{t=1}^{N-1} (x_{t+1} - \bar{x}_{(2)})^2}} \quad (2.41)$$

Por analogía, con la ecuación (2.41) se tiene que la media de las primeras  $(N - 1)$  observaciones, será:

$$\bar{x}_{(1)} = \sum_{t=1}^{N-1} \frac{x_t}{(N - 1)} \quad (2.42)$$

Mientras que la media de las últimas  $(N - 1)$  observaciones estará dado por:

$$\bar{x}_{(2)} = \sum_{t=2}^N \frac{x_t}{(N - 1)} \quad (2.43)$$

Debido a que el coeficiente dado por la ecuación (2.41) mide la correlación entre observaciones sucesivas, éste es el conocido como el coeficiente de autocorrelación o coeficiente de correlación en serie.

Si  $\bar{x}_{(1)} @ \bar{x}_{(2)}$ , la ecuación (2.41) queda de la siguiente manera:



$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{(N-1) \sum_{t=1}^N (x_t - \bar{x})^2 / N} \quad (2.44)$$

Donde:  $\bar{x} = \frac{\sum_{t=1}^N x_t}{N}$

Debido a que muchos autores desechan el factor  $N/(N-1)$  que es aproximadamente igual a uno (1) para  $N$  grande, se simplifica la formula tomando eso en consideración:

$$r_1 = \frac{\sum_{t=1}^{N-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2} \quad (2.45)$$

Ahora bien, para hallar la correlación entre observaciones separadas a una distancia  $k$ , se puede hacer de una manera similar a como se ha trabajado, es decir, dado por:

$$r_k = \frac{\sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^N (x_t - \bar{x})^2} \quad (2.46)$$

Donde  $r_k$  es conocido como el coeficiente de autocorrelación de retraso  $k$ .



Por otra parte, se debe considerar que en la práctica es muy común ver que los coeficientes de autocorrelación se calculen por medio de la serie de coeficiente de autocovarianza,  $\{C_k\}$ , entonces al definir por analogía con la formula usual de autocovarianza se tiene:

$$C_k = \frac{1}{N} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \quad (2.47)$$

Por lo cual se puede calcular:

$$r_k = \frac{C_k}{C_0} \quad (2.48)$$

Definido para  $k = 1, 2, \dots, m$ , donde  $m < N$ . Ahí está usualmente un pequeño punto en calcular  $r_k$  para valores de  $k$  mayores que  $N/4$ . Ahora, algunos autores sugieren:

$$C_k = \frac{1}{N-k} \sum_{t=1}^{N-k} (x_t - \bar{x})(x_{t+k} - \bar{x}) \quad (2.49)$$

Que es similar a la ecuación (2.47) pero con una pequeña diferencia para  $N$  grande.

#### 2.1.7.1. DETECCIÓN DE AUTOCORRELACIÓN.



Arranz y Zamora (2000), mencionan en su libro que para detectar la presencia de autocorrelación, se pueden utilizar métodos gráficos así como también representa una buena opción el aplicar contrastes de hipótesis. Al usar los métodos gráficos se debe intuir si existe autocorrelación cuando existan comportamientos sistemáticos para los residuos. Mientras que en los contrastes de hipótesis, se permite, mediante el uso de una regla de decisión, considerar si con los datos de la muestra y con un nivel concreto de significación ( $\alpha$ ) se debe o no rechazar la hipótesis nula.

Todos los contrastes numéricos de autocorrelación se plantean con idénticas hipótesis; de modo que puede señalar la hipótesis nula y la alternativa del contraste como sigue:

$H_0$ : No existe autocorrelación.

$H_1$ : Existe autocorrelación.

Desde otro punto de vista, se puede decir que en la hipótesis nula se considera que el término de perturbación correspondiente a una observación, es independiente del correspondiente a cualquier otra observación. Por el contrario, en la hipótesis alternativa se considera que el término de error del modelo econométrico se encuentra autocorrelacionado a través del tiempo, en dicha hipótesis al considerar la existencia de un patrón de comportamiento





para los residuos, se pueden clasificar bajo un esquema de procesos autorregresivos  $AR(p)$ , de medias móviles  $MA(q)$  o mixtos  $ARMA(p,q)$ , dependiendo del contraste que se vaya a utilizar.

A continuación se presentan dos contrastes que permiten detectar si las perturbaciones se encuentran autocorrelacionadas o no, que en caso de estarlo, se señalará bajo qué esquema.

#### 2.1.7.1.1. CONTRASTE $d$ DE DURBIN-WATSON.

Para detectar la presencia de autocorrelación en una serie de datos, la prueba frecuentemente utilizada, es el contraste desarrollado por Durbin y Watson (1951). En este, se verifica la hipótesis de que no existe autocorrelación frente a la alternativa de que si existe autocorrelación de primer orden bajo un esquema autorregresivo:

$$AR(1): u_t = \rho u_{t-1} + \varepsilon_t$$

(2.50)

A continuación se explica el contraste de Durbin y Watson, comenzando con la formulación de la hipótesis:

$$H_0 : \rho = 0 \quad \text{No existe autocorrelación } AR(1)$$



$$H_1: 0 < |\rho| < 1 \quad \text{Existe autocorrelación } AR(1)$$

La forma concreta de la hipótesis alternativa establece unas cotas para el coeficiente de correlación; las cuales son necesarias para garantizar algunas características del modelo, específicamente que la varianza es finita y se trata por tanto de un proceso no explosivo. El estadístico de prueba se define de la siguiente manera:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} \quad (2.51)$$

Partiendo de la ecuación (2.46) se puede decir que:

1. Si hay presencia de autocorrelación positiva, las diferencias entre residuos que distan un periodo es muy pequeño y por tanto, el valor del estadístico  $d$  será próximo a cero.
2. Si hay presencia de autocorrelación negativa, los residuos serán aproximadamente iguales pero con signos contrarios, por lo tanto la diferencia entre ellos será grande y el estadístico se encontrará más



cercano al límite superior, el cual será establecido con un valor de cuatro como se verá más adelante.

3. Si no hay presencia de autocorrelación, la relación entre los residuos será intermedia y por tanto, el valor del estadístico experimental también alcanzará un valor intermedio.

Los límites de variación del estadístico  $d$ , se pueden obtener a partir del desarrollo de la ecuación (2.51), de la cual resulta una expresión en función del coeficiente de autocorrelación muestral de primer orden para los residuos  $\hat{\rho}$ , tal como sigue:

$$d = \frac{\sum_{t=2}^n (e_t - e_{t-1})^2}{\sum_{t=1}^n e_t^2} = \frac{\sum_{t=2}^n (e_t^2 + e_{t-1}^2 - 2e_t e_{t-1})}{\sum_{t=1}^n e_t^2} = \frac{\sum_{t=2}^n e_t^2 + \sum_{t=2}^n e_{t-1}^2 - 2 \sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \quad (2.52)$$

Cuando el tamaño de la muestra es grande, se puede considerar lo siguiente:

$$\sum_{t=2}^n e_t^2 \gg \sum_{t=2}^n e_{t-1}^2 \gg \sum_{t=1}^n e_t^2, \text{ por lo que el estadístico } d \text{ se puede expresar}$$

como sigue:

$$d \gg \frac{2 \sum_{t=2}^n e_t^2 + 2 \sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \quad (2.53)$$



Por otro lado, dado que el coeficiente de correlación empírico de primer orden se calcula mediante el uso de  $\hat{\rho}$ , se llega a la expresión:

$$\hat{\rho} = \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=1}^n e_t^2} \quad (2.54)$$

Por lo tanto, el estadístico  $d$  se puede expresar como:  $d = 2(1 - \hat{\rho})$

Donde, los límites de variación del coeficiente de correlación empírico estarán dados por:  $-1 \leq \hat{\rho} \leq 1$ . Por lo que es posible deducir el rango de variación que puede tomar el estadístico de Dubin-Watson en conjunto con el signo de la autocorrelación de la siguiente forma:

Si  $\hat{\rho} = -1 \Rightarrow d = 4$  Se puede afirmar que existe autocorrelación negativa.

Si  $\hat{\rho} = 0 \Rightarrow d = 2$  Indica la ausencia de autocorrelación.

Si  $\hat{\rho} = 1 \Rightarrow d = 0$  Se puede afirmar que existe autocorrelación positiva.

De ese modo se puede observar que el estadístico experimental tomará valores que estarán ubicados entre 0 y 4, lo que significa que mientras el valor del estadístico  $d$  se aproxime a cero o a cuatro, mayor será la evidencia de la presencia de autocorrelación positiva o negativa según sea el



caso. En el otro escenario, si el valor del estadístico  $d$  es dos, quiere decir que la correlación muestral será nula y por tanto no se detectará un problema de autocorrelación entre las perturbaciones.

Es importante resaltar que los límites extremos (con valores de  $0,2$  y  $4$ ), deben matizarse estableciendo regiones más amplias en las que pueda considerarse si existe o no autocorrelación y, en caso de detectarse, determinar si es positiva o negativa. En este sentido es preciso aclarar que no es sencilla la distribución teórica de este estadístico y depende de los valores concretos de la matriz de regresores; por tal motivo, no se puede establecer una regla de decisión con un valor crítico único.

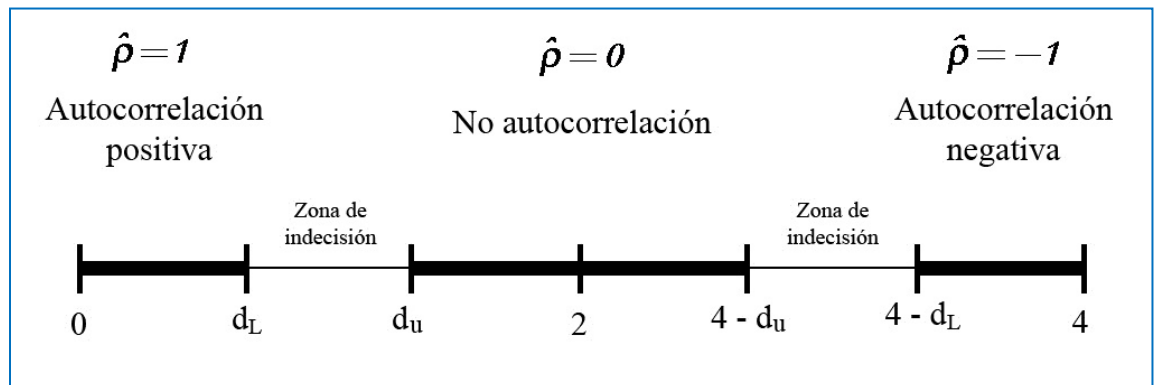
Con el fin de enfrentar y solucionar el problema antes mencionado, Durbin y Watson hallaron unos valores que permiten tomar decisiones acerca de la presencia o ausencia de autocorrelación, que son conocidos como el límite superior  $d_U$  e inferior  $d_L$ .

Estos valores señalan el límite superior  $d_U$  para considerar autocorrelación positiva, esto es, para valores del estadístico experimental superiores a este límite no se rechaza la hipótesis de ausencia de autocorrelación y el límite inferior  $d_L$  para no rechazar la hipótesis nula y suponer que las covarianzas de las perturbaciones del modelo son nulas y, por lo tanto, no están autocorrelacionadas



Si el valor del estadístico  $d$  es superior a dos, se puede contrastar la hipótesis nula de no autocorrelación frente a la alternativa de autocorrelación negativa. El análisis es similar pero considerando el valor máximo de 4 como límite para la autocorrelación negativa, por lo tanto, los límites anteriores se establecen en los puntos  $4 - d_U$  y  $4 - d_L$ .

Con el fin de aclarar lo antes expuesto, a continuación se mostrará gráficamente las regiones del contraste de Durbin y Watson:



**Figura 2.2.** Regiones del estadístico de Durbin y Watson.

**Fuente:** Durbin y Watson (1950).

Escrito de otra forma:

**Tabla 2.3.** Reglas de decisión del contraste Durbin Watson.

Regla	Decisión
-------	----------



$0 < d < d_L$	Se rechaza $H_0$ , existe autocorrelación positiva con un esquema $AR(1)$ .
$d_L < d < d_U$	El contraste no es concluyente.
$d_U < d < 4 - d_U$	No se rechaza $H_0$ , no existe autocorrelación.
$4 - d_U < d < 4 - d_L$	El contraste no es concluyente.
$4 - d_L < d < 4$	Se rechaza $H_0$ , existe autocorrelación negativa con un esquema $AR(1)$ .

**Fuente:** Durbin, J. and Watson (1950).

Donde los límites dependen del tamaño de la muestra ( $n$ ) y del número de regresores del modelo ( $k$ ). Ver tabla A.1 en el apéndice I.

Este contraste requiere el tratamiento empírico con las siguientes fases:

1. Estimación por mínimos cuadrados ordinarios (MCO) del modelo de regresión.
2. Cálculo de los residuos MCO.
3. Obtención del estadístico  $d$  (experimental) de Durbin-Watson.
4. Búsqueda de los niveles críticos del contraste.
5. Aplicación de la regla de decisión.



El contraste de Durbin–Watson presenta un inconveniente: a veces no puede ser concluyente, por lo que hay que utilizar otros criterios para determinar si existe o no autocorrelación. En este sentido una solución clásica consiste en ampliar las regiones de rechazo, considerando de ese modo que existe autocorrelación positiva para valores del estadístico  $d$  inferiores a  $d_U$  y autocorrelación negativa para los valores del estadístico experimental superiores a  $4 - d_U$ .

Es necesario reflexionar en el hecho de que este estadístico se basa en un conjunto de supuestos, por lo que se puede mencionar lo siguiente:

En primer lugar hay que señalar que el diseño original del contraste se basó en el análisis de un modelo de regresión que incluía término independiente. Sin embargo, este requisito exigible al modelo fue posteriormente resuelto. Farebrother (1980), calculó los valores críticos del contraste para los modelos en los que no existe término independiente.

Junto con la necesidad de término independiente en el modelo, es también un requisito que la matriz de variables explicativas no sea aleatoria, es decir, debe ser determinista y fija en un muestreo repetido. Por tanto, este contraste no es válido en modelos dinámicos que consideren como regresor retardos de la variable dependiente.

La hipótesis alternativa considera que, si las perturbaciones están autocorrelacionadas, el proceso que las genera es autorregresivo de orden 1;





esto es:  $u_t = \sigma u_{t+1} + \varepsilon_t$ . No obstante, se ha comprobado que este estadístico es robusto frente a otras especificaciones de la hipótesis alternativa y, además permite detectar errores de especificación derivados de falta de especificación dinámica y/o de la omisión de variables que estén correlacionadas.

#### 2.1.7.1.2. ESTADÍSTICO $h$ DE DURBIN.

Anteriormente se estudio el contraste de Durbin y Watson, que impone como condición para su correcta interpretación que los modelos contengan regresores exclusivamente no aleatorios; por lo que no se puede aplicar en modelos dinámicos en los que se considere como regresor algún retardo de la variable dependiente. Para corregir esta deficiencia, Durbin y Watson (1971), desarrolló un estadístico que sí puede aplicarse en los modelos que incluyan retardos de la variable dependiente. Para este caso se ha obtenido un test asintótico para muestras grandes.

La formulación de la hipótesis nula sigue siendo la misma ya que es un contraste para la autocorrelación de orden uno bajo un esquema autorregresivo

$$AR(1): u_t = \rho u_{t+1} + \varepsilon_t$$

$$H_0: \rho = 0 \text{ No existe autocorrelación } AR(1)$$

La hipótesis alternativa, por su parte, se especifica ahora de modo que el contraste se configure como un contraste unilateral; es decir, se van a



establecer dos posibles hipótesis alternativas según se considere que la autocorrelación puede ser positiva o negativa. Así, el contraste quedaría especificado como:

$$\begin{array}{ll} H_0 : \rho = 0 & H_0 : \rho = 0 \\ H_1 : \rho < 0 & H_1 : \rho > 0 \end{array}$$

El estadístico de prueba se puede expresar como:

$$h = \hat{\rho} \sqrt{\frac{n}{1 - n[\text{Var}(b_i)]}} \quad (2.55)$$

El cual se distribuye asintóticamente según una distribución  $N(0,1)$ , con un nivel de significación del 5% con el cual se supone no rechazar la hipótesis nula para los valores de  $h$  pertenecientes al intervalo  $(-1.645; 1.645)$ , ya que se trabaja con un contraste de una sola cola.

Para el cálculo de ese estadístico se necesitan conocer los siguientes datos:

- Tamaño de la muestra  $(n)$ .
- Varianza muestral estimada del coeficiente del regresor aleatorio  $(Y_{t-1})$  en la regresión MCO del modelo a estimar;



es decir, obtenida bajo el supuesto del Modelo de Regresión Lineal Normal Clásico:  $MRLNC[Var(b_i)]$ .

- Coeficiente de correlación estimado ( $\hat{\rho}$ ).

El coeficiente de correlación estimado se puede calcular a partir de la estimación de una estructura autorregresiva de orden 1 para los residuos, una regresión MCO de los residuos frente a un retardo de los mismos ( $e_t = \rho e_{t-1} + \varepsilon_t$ ), esto es:

$$\hat{\rho} \approx \frac{\sum_{t=2}^n e_t e_{t-1}}{\sum_{t=2}^n e_{t-1}^2} \quad (2.56)$$

Otra opción puede ser el calcular esta correlación muestral a partir del valor del estadístico de prueba del contraste de Durbin-Watson:

$$\hat{\rho} \approx 1 - \frac{d}{2} \quad (2.57)$$

Este contraste requiere de la realización de las siguientes fases:

- Estimación MCO del modelo de regresión y obtención de la varianza estimada del coeficiente del regresor aleatorio  $[Var(b_i)]$ .
- Cálculo del coeficiente de correlación estimado.
- Cálculo del valor del estadístico experimental  $h$ .



- Aplicación de la siguiente regla de decisión: Si  $h > 1,645$  se rechaza la hipótesis nula con un nivel de significación del 5% considerando entonces que existe autocorrelación negativa de
- primer orden, el valor del estadístico experimental  $h$  debe ser inferior a  $-1,645$ .

Este contraste presenta un inconveniente muy importante: que si el radicando es negativo la prueba falla, es decir, cuando  $[n\text{Var}(b_i) > 1]$ .

### 2.1.8. GRÁFICOS DE CONTROL.

#### 2.1.8.1. GRÁFICOS DE CONTROL DE SHEWHART.

Según Montgomery (2005), los gráficos para el control de productos industriales fueron desarrollados inicialmente por W. Shewhart en 1931, con el principal objetivo de investigar si un proceso se encuentra bajo control estadístico. El elemento clave en los gráficos de control es la muestra de control, que sirve para construir el gráfico y monitorizar el estado del procedimiento analítico. Esta muestra, que tiene que ser estable con el tiempo, puede ser:

- Una sustancia patrón.



- Una muestra sintética adicionada.
- Un material de referencia o un material de referencia certificado.
- Una muestra real.

En la mayoría de estos tipos de muestras el valor de la concentración o propiedad que se desea monitorizar nos viene dado (en las sustancias patrón, materiales de referencia o materiales de referencia certificados), o bien se conoce de una forma muy exacta (en el caso de muestras sintéticas fortificadas). Pero en el tipo de muestras de control más utilizado (una muestra real), se desconoce este valor de la concentración o propiedad a controlar. En este tipo de muestras la estimación de la concentración o propiedad a monitorizar se debe llevar a cabo analizando la muestra de control con nuestro método analítico una vez acabado de verificar la trazabilidad del mismo.

El fundamento de los gráficos de control se basa en la asunción de la normalidad de los resultados de medida: cuando se lleva a cabo algún proceso (por ejemplo, un método de análisis) de forma sistemática, es decir, bajo las mismas fuentes de influencia o variación, el proceso se verá afectado por errores aleatorios que conducirán a una distribución normal de los resultados. Esta afirmación es una consecuencia del teorema del límite central. Se dirá que el método analítico está bajo control si los resultados obtenidos



con este método siguen las características de una distribución normal. Por ejemplo, aproximadamente el 67% de los resultados han de encontrarse dentro del intervalo  $1\sigma$  (donde  $\sigma$  es desviación estándar asociada a los análisis de la muestra de control con el procedimiento que se desea monitorizar), aproximadamente el 95% de los resultados han de encontrarse dentro del intervalo:  $2\sigma$  y aproximadamente el 99% de los resultados han de encontrarse dentro del intervalo:  $3\sigma$ . Cuando los resultados de los análisis de la muestra de control a lo largo del tiempo se encuentran dentro de los límites aceptados, se dice que el sistema se encuentra bajo control estadístico. Cuando se encuentran puntos fuera de los límites especificados, o se encuentran tendencias, se dice que el sistema se encuentra fuera de control.

En la construcción de un gráfico de control se pueden distinguir las siguientes etapas:

- Etapa de aprendizaje. En esta etapa se obtienen los resultados iniciales con la muestra de control. En el caso de utilizar una muestra real, se debería comprobar la normalidad y la presencia de resultados discrepantes y su eliminación. Con los resultados iniciales de la muestra de control se establece el valor de la línea central. Este valor debería obtenerse con un mínimo de 15-30 análisis de la muestra de control. Los diferentes límites suelen establecerse a una distancia del valor



central  $2 \sigma$  (línea de aviso), y a una distancia del valor central  $3 \sigma$  (línea de control).

Los límites de aviso y de control situados a unas distancias de  $2 \sigma$  y  $3 \sigma$  respectivamente, pueden construirse utilizando los valores  $2$  y  $3$  cuando el valor promedio de la muestra de control ha sido encontrado con un número suficientemente grande de repeticiones (alrededor de  $30$ ). En este caso se asume que se conocen los valores reales de los parámetros (promedio y desviación estándar). Si se tienen menos repeticiones, se aconseja considerar que los valores reales de estos parámetros son desconocidos, y se deben efectuar correcciones sobre la asunción de distribución normal. Esto implica utilizar valores tabulados, (Grant 1988) en lugar de los valores  $2$  y  $3$ . Normalmente un laboratorio empieza considerando como desconocidos los valores de los parámetros, hasta que se han recogido suficientes datos como para poder considerar estos parámetros como conocidos.

- Etapa de control. En esta etapa se representan frente al tiempo los diversos resultados de la muestra de control con el objetivo de detectar tendencias y situaciones fuera de control.

El gráfico de control se denomina un gráfico de valores individuales, y en ella se efectúa un solo replicado de la muestra de control cada vez que ésta se analiza.



El gráfico de valores individuales es quizá el más empleado, pero hay otras variaciones:

- Gráfico de promedios: cada vez que se analiza la muestra de control se efectúan  $n$  réplicas (siempre el mismo número  $n$  de réplicas), y se representa el valor medio de estas  $n$  réplicas. En este caso, los límites de aviso y de control se encuentran respectivamente a  $2\sigma/\sqrt{n}$  y  $3\sigma/\sqrt{n}$ , siempre que se consideren como conocidos los parámetros del valor promedio y de la desviación estándar. En estos gráficos se aplica el mismo comentario para la construcción de los límites de aviso y control que en el gráfico de valores individuales cuando se tienen pocos datos y no se pueden considerar los parámetros promedio y desviación estándar como verdaderos.
- Gráfico de intervalos móviles: se mide el intervalo entre un par consecutivo de valores en un gráfico de valores individuales. De esta manera se pueden controlar ‘saltos’ en el procedimiento analítico
- Gráfico de intervalos: cuando se establece el control de los promedios, su distribución puede medirse por el intervalo (valor





máximo – valor mínimo) de los valores utilizados para cada promedio, y controlar así esta dispersión.

### 2.1.8.2. GRÁFICOS DE CONTROL CUSUM.

Page (1954), propuso los gráficos de control CUSUM (Sumas Acumulativas) para la detección de cambios menores a dos veces sigma, el cual a través de simulaciones concluyó que estos gráficos son más eficientes que los de Shewhart para detectar pequeños cambios en la media del proceso.

**Tabla 2.4.** Porcentaje de señales diagnosticadas en cambios de media y varianza para cambios en la muestra 50 (APC).

$\delta$	Parámetro	Porcentaje de significancia nominal en pruebas de seguimiento						
		$\sigma$						
		0.51	0.64	0.8	1	1.25	1.56	1.95
0	Media	8	25	66	82	77	40	13
	Varianza	100	99	94	90	86	88	90
0.5	Media	65	71	84	88	82	48	18
	Varianza	95	82	61	57	64	77	84
1	Media	99	99	99	97	89	62	28
	Varianza	66	43	23	14	21	49	70
1.5	Media	100	100	99	96	85	62	35



	<b>Varianza</b>	34	21	12	10	14	33	56
<b>2</b>	<b>Media</b>	99	98	96	89	77	56	33
	<b>Varianza</b>	17	12	8	8	12	27	48

**Fuente:** Hawkins (2005b)

Hawkins y Olwel (1997) plantean el gráfico CUSUM para el control de la media del proceso de la siguiente forma:

Suponiendo que la varianza del proceso  $\sigma^2$  es constante y que se quiere monitorear la media  $\mu$  de una característica de calidad  $X$ , la cual sigue una distribución  $N(\mu, \sigma^2)$ . Se tiene un valor objetivo  $\mu_0$  para  $\mu$ , del cual se quieren detectar desviaciones significantes. Si el valor estandarizado de  $X$  es  $Z_i = (X_i - \mu_0) / \sigma$ , donde  $\sigma$  es la desviación estándar de  $X_i$  o de  $\bar{X}_i$ , una carta de sumas acumuladas de estos valores contra el número total de muestras recolectadas. Para este procedimiento se calculan las siguientes sumas:

$$\begin{aligned} CUM^+ &= C_{i-1}^+ + (Y_i - \mu) - k \\ CUM^- &= C_{i-1}^- + (Y_i - \mu) + k \end{aligned} \tag{2.58}$$

Donde:

$k$ : un valor de referencia o de tolerancia constante.



La elección del valor  $k$  deberá hacerse respecto de la magnitud del corrimiento ( $\delta$ ) que quiera detectarse. En Montgomery (2005) se asegura que con frecuencia se escoge aproximadamente a la mitad entre el objetivo  $\mu_0$  y el valor considerado fuera de control de la media  $\mu_1$ , el cual está interesado en detectar con rapidez el analista. Por tanto, si el cambio se expresa en unidades de desviación estándar como  $\mu_1 = \mu_0 + \delta\sigma$ , entonces  $k$  es la mitad de la magnitud del corrimiento y queda expresado de la siguiente forma:

$$k = \frac{\delta}{2}\sigma = \frac{|\mu_1 - \mu_0|}{2} \tag{2.59}$$

Los gráficos de CUSUM se basan en las siguientes ecuaciones:

$$\begin{aligned} C_i^+ &= \max\left\{\frac{2h\sigma}{3}, CUM^+\right\} \\ C_i^- &= \max\left\{\frac{2h\sigma}{3}, CUM^-\right\} \end{aligned} \tag{2.60}$$

Donde:

$h$ : es el intervalo de decisión.

Entonces se puede expresar que el proceso está fuera de control cuando:



$$\begin{array}{l} \vdots \\ \vdots \\ \vdots \\ C_i^- \geq \frac{h\sigma}{2} \\ \vdots \\ \vdots \\ \vdots \\ C_i^+ \leq \frac{h\sigma}{2} \\ \vdots \\ \vdots \\ \vdots \end{array} \quad (2.61)$$

La selección correcta de los parámetros  $k$  y  $h$  es muy importante, ya que tienen un impacto sustancial sobre el desempeño del gráfico CUSUM. En Montgomery (2005) se ofrecen recomendaciones para seleccionarlos y en general proponen usar un valor de  $h$  igual a  $4\sigma$  o  $5\sigma$ , donde en el caso de que  $k = 1/2$  por lo general se obtendrá una CUSUM que tiene propiedades convenientes de la ARL contra un corrimiento de aproximadamente  $1\sigma$  en la media del proceso.

Una vez construido el gráfico de control CUSUM, Montgomery (2005) señala en su libro, que se puede indicar el momento probable en el cual ocurrió el corrimiento de la media, a través del uso de dos contadores  $N^+$  y  $N^-$ , que a la posición donde el gráfico detecta el cambio se le resta el valor del contador según sea el caso, es decir, si se detecta el cambio por el lado superior  $C_i^+$  se usa  $N^+$  y de ser detectado el cambio por el lado superior  $C_i^-$  se usa  $N^-$ . El contador  $N^+$  registra el número de periodos consecutivos desde que la CUSUM del lado superior  $C_i^+$  subió por encima del valor cero, mientras que el contador  $N^-$  registra el número de periodos consecutivos desde que la CUSUM del lado inferior es menor que cero.



Hawkins (2005b) realizó una comparación de los métodos del APC y el CUSUM, determinando que para cambios menores a 1,5 y superiores a 2, el grafico CUSUM tiene mejor capacidad de detección en la media, además que el CUSUM detecta mejor los aumentos en la variabilidad, mientras que el APC los detecta mejor en disminución de la variabilidad.

### 2.1.8.3. GRÁFICOS DE CONTROL EWMA.

Los gráficos de control EWMA (Promedio Móvil Ponderado Exponencialmente) fueron propuestos por Roberts (1959), en el cual, para la detección de cambios menores a dos veces sigma, a través de simulaciones concluyó que estos gráficos son más eficientes que los de Shewhart, representando al igual que el gráfico CUSUM una buena alternativa para detectar pequeños cambios en la media del proceso. Así mismo, estos demostraron ser robustos frente a la no normalidad de los datos en simulaciones realizadas por Borrór, Montgomery y Runger (1999), donde compara la longitud de corrida promedio (ARL) de la carta de control Shewhart para mediciones individuales con la carta de control EWMA para casos no normales, usando distribuciones Gamma para representar casos de distribuciones sesgadas, y distribuciones  $t$  para distribuciones simétricas con colas más cargadas; de igual manera, Peña y Quiroz (2009) indican una comparación de desempeño de los gráficos EWMA, CUSUM y Shewhart,



mostrando los gráficos EWMA superioridad con respecto a los cambios de distribución.

El Promedio Móvil Ponderado Exponencialmente se define como:

$$Z_t = \lambda X_t + (1 - \lambda)Z_{t-1} \quad (2.62)$$

Donde  $X_t$  es el valor observado individualmente en el tiempo del proceso o también puede ser el promedio de las muestras tomadas en el tiempo cuando el tamaño del subgrupo racional es mayor a uno.  $\lambda$  es una constante de suavizamiento que determina la memoria del estadístico EWMA y su valor está entre  $0 < \lambda \leq 1$ . El valor inicial  $Z_0$  requerido en la primera muestra ( $i = 1$ ) es el objetivo del proceso, de modo que  $Z_0 = \mu_0$ .

Los límites de control y la línea central para el gráfico de control EWMA son los siguientes respectivamente:

$$LCS = \mu_0 + L \frac{\sigma}{\sqrt{n}} \sqrt{\frac{\lambda}{2 - \lambda}} \quad (2.63)$$

$$LCI = \mu_0 - L \frac{\sigma}{\sqrt{n}} \sqrt{\frac{\lambda}{2 - \lambda}} \quad (2.64)$$

$$LC = \mu_0 \quad (2.65)$$



Un punto muy importante a tratar es la elección de  $\lambda$ , la cual el analista debe elegir según su criterio. Algunas de las formas para determinar dicho valor se mencionan a continuación:

- Según Hunter (1986) se recomienda el uso de:  $\lambda = 0.2$  y  $0.1$ .
- Calcular  $\lambda$  usando un proceso iterativo de mínimos cuadrados como sigue:

Para diferentes valores de  $\lambda$  se calcula  $\sum_{i=0}^t e_i^2$  y se escoge el valor  $\lambda$  de que minimice la suma de cuadrados, en el cual:

$$e_t = X_t - Z_t$$

(2.66)

Donde:

$X_t$  es el valor individual observado o la media.

$Z_t$  es el EWMA estimado.

Cálculo en forma conjunta de los valores de  $\lambda$  y  $L$ .

Existen diferentes formas, las cuales se mencionan a continuación:



- En la publicación de Aparisi y García-Díaz (2004), se encuentran los valores óptimos para los siguientes casos:

- Valor del ARL bajo control que desee que tenga el gráfico.
- El tamaño de la muestra.
- El tamaño de cambio para el cual se optimiza el gráfico, es decir, se busca el EWMA más rápido para detectar ese tamaño de cambio, introducido en unidades de sigma.

- La estrategia del diseño está basada en seleccionar  $(\lambda, L)$  de forma tal que dada la longitud media de corrida (LMC) del diagrama de un cambio  $D$  en la media, sea mínima. La combinación  $(\lambda, L)$  es óptima en el sentido de que para la elección de una LMC bajo control, producirá una LMC fuera de control mínima. Crowder (1989), recomienda las siguientes etapas, para un diseño óptimo del gráfico EWMA:

Etapas I:

Elegir la menor LMC aceptable cuando el cambio en la media es  $D = 0$ . La selección de la LMC está basada informalmente sobre condiciones, tales como el costo asociado a una falsa alarma y el tiempo del proceso.

Etapas II:





Decidir que magnitud del cambio en la media debe ser detectada lo más rápido posible.

Etapa III:

Una vez que se tenga la LMC de la etapa I y la magnitud del cambio de la etapa II, en la tabla correspondiente a la LMC de la etapa I, se escogen los valores de  $\lambda$  y  $k$  que hacen mínima la LMC para el cambio especificado en la etapa II.

Crowder (1987), muestran que la LMC del diagrama EWMA puede ser expresada como la solución de la ecuación integral, que a través de unas conversiones queda:

$$\begin{aligned}
 L(d) &= 1 + \frac{1}{\lambda} \sum_{i=1}^m W_i L(P_i) f\{[P_i - (1 - \lambda)d]/\lambda\} \\
 L(P_i) &= 1 + \frac{1}{\lambda} \sum_{i=1}^m W_i L(P_i) f\{[P_i - (1 - \lambda)P_i]/\lambda\} \\
 L(P_m) &= 1 + \frac{1}{\lambda} \sum_{i=1}^m W_i L(P_i) f\{[P_i - (1 - \lambda)P_m]/\lambda\} \quad (2.67)
 \end{aligned}$$

Asignándoles valores a  $\lambda$  y  $L$ , el sistema de ecuaciones lineales, se puede resolver y obtener la LMC del cambio  $D = (L(d))$ . Se elaboraron tablas para LMC bajo control de 100, 250, 370, 500, resolviendo el sistema de ecuaciones lineales con ayuda de un programa en FORTRAN, utilizando



cuadratura de Gauss de 24 puntos. Ver Tabla 2 en apéndice I que contiene los valores de  $\lambda$  y  $L$  para diferentes valores de “ $d$ ”.

### **2.1.9. ANÁLISIS DE PUNTO DE CAMBIO (APC).**

Según Gel (2009), los puntos de cambio son aquellos donde dos segmentos adyacentes de una serie de tiempo están conectados; estos ocurren como un cambio en la media, la varianza, en algún parámetro, o en el modelo estructural, como también en cualquiera de sus combinaciones. Por lo tanto el APC, se puede establecer como el análisis realizado a una serie de tiempo, en el cual se estudia el punto donde ocurre el cambio.

#### **2.1.9.1. MODELO BAYESIANO PARA ANÁLISIS DE PUNTO DE CAMBIO.**

Hawkins y Zamba (2005b), proponen una formulación del análisis de punto de cambio, por medio de una prueba de hipótesis. En el cual se ofrece un solo diagnóstico para detectar un cambio sostenido en la media, la varianza, o en ambos, en lugar de dos diagnósticos distintos. En dicho enfoque, se utiliza una prueba de razón de verosimilitud (MLE) para detectar un cambio en datos que siguen una distribución normal. Usando ésta formulación, no se asume que los parámetros son conocidos con exactitud (lo que si se hace con los gráficos de Shewhart, CUSUM, EWMA), reconociendo de ésta forma la realidad, que en el mejor de los casos se tiene una



estimación de parámetros razonable y no sus valores exactos. A continuación, se expone la propuesta.

Se supone que mientras el proceso está bajo control, las lecturas siguen una distribución normal y son estadísticamente independientes, la media, la varianza, o ambos pueden cambiar cuando el proceso cruza el punto de cambio  $\tau$ . Entonces, modelando las lecturas del proceso por dos distribuciones normales se tiene:

$$X_i \sim \begin{cases} N(\mu_1, \sigma_1^2) & \text{si } i \leq \tau \\ N(\mu_2, \sigma_2^2) & \text{si } i > \tau \end{cases} \quad (2.68)$$

Donde  $X_1, X_2, \dots, X_n$  representan las lecturas del proceso sucesivas. La distribución bajo control es  $N(\mu_1, \sigma_1^2)$ , es decir, las lecturas seguirán esta distribución hasta algún instante  $t$  (si lo hay), que no es más que el punto de cambio en el cual la distribución normal cambia de ubicación si se produce que  $\mu_1 \neq \mu_2$ , y cambia de escala si se produce  $\sigma_1 \neq \sigma_2$ . Uno o ambos de estos cambios podrían ocurrir.

Si las  $n$  lecturas del proceso se han acumulado y, para  $0 \leq i \leq k \leq n$ , definiendo el resumen de las estadísticas se tiene:

$$\bar{X}_{i,k} = \frac{1}{k-i} \sum_{j=i+1}^k X_j$$



$$(2.69)$$

$$V_{i,k} = \sum_{j=i+1}^k (X_j - \bar{X}_{i,k})^2 \quad (2.70)$$

Hawkins y Zamba (2005b) mencionan en su publicación que se pueden calcular las estimaciones convencionales de los parámetros. Suponiendo que se conoce que el punto de cambio se encuentra en el instante  $\tau = k$  existen varios escenarios que se pueden presentar:

- En los casos donde el cambio detectado ocurre en la media del proceso, en la varianza o ambos, se usan las siguientes ecuaciones según sea el caso:

$$\hat{\mu}_1 = \bar{X}_{0,k} \quad (2.71)$$

$$\hat{\mu}_2 = \bar{X}_{k,n} \quad (2.72)$$

$$\bar{\sigma}_1^2 = (V_{0,k}) / (k - 1) \quad (2.73)$$

$$\bar{\sigma}_2^2 = (V_{k,n}) / (n - k - 1) \quad (2.74)$$

- En el caso de detectar el cambio únicamente en la varianza, la estimación de la media (constante) sería:

$$\hat{\mu} = \bar{X}_{0,n} \quad (2.75)$$



- En el caso de detectar el cambio únicamente en la media, la estimación de la varianza (constante) sería:

$$\bar{\sigma}^2 = (V_{0,k} + V_{k,n}) / (n - 2) \quad (2.76)$$

Es evidente que es necesaria la determinación de si hubo un cambio y donde ocurrió aproximadamente para poder calcular los parámetros antes y después del cambio (si lo hay).

Para comenzar, si el punto de cambio ocurre en el instante  $\tau = k$ , entonces la prueba estadística de razón de verosimilitud generalizada (GLR) se puede expresar como:

$$GLR = k \log \frac{S_{0,n}}{S_{0,k}} + (n - k) \log \frac{S_{0,n}}{S_{k,n}} \quad (2.77)$$

En donde se define  $S_{ij} = V_{ij} / (j - i)$ , será el MLE de la variación de la secuencia de  $X_{i+1}, \dots, X_n$  (sin los grados de libertad habituales de ajuste de sesgo del denominador).

En la hipótesis nula de que no ocurra ningún cambio, este estadístico tiene una distribución asintótica chi-cuadrado con 2 grados de libertad. Lawley



(1956) menciona que la calidad de esta aproximación se puede mejorar sustancialmente haciendo la corrección de Bartlett, en el cual se divide por un factor que hace que la esperanza del GLR sea igual a los grados de libertad. La esperanza del GLR se conoce pero implica la función digamma, una de las funciones trascendentes menos familiares. Una expansión estándar de la confianza muestra a los términos de orden  $O(n^{-2})$

$$E[GLR] = 2 + \frac{11}{6} \frac{\psi(1)}{k} + \frac{1}{n-k} - \frac{1}{n} \frac{\psi(1)}{k^2} + \frac{1}{(n-k)^2} - \frac{1}{n^2} \frac{\psi(1)}{k} \quad (2.78)$$

Esto conlleva a la prueba estadística de Bartlett corregida:

$$G_{k,n} = \frac{C \left[ k \log \frac{S_{0,n}}{S_{0,k}} + (n-k) \log \frac{S_{0,n}}{S_{k,n}} \right]}{C}$$

$$C = 1 + \frac{11}{12} \frac{\psi(1)}{k} + \frac{1}{n-k} - \frac{1}{n} \frac{\psi(1)}{k^2} + \frac{1}{(n-k)^2} - \frac{1}{n^2} \frac{\psi(1)}{k} \quad (2.79)$$

Donde  $C$  es el factor de corrección de Bartlett. Si el punto de cambio no se conoce a priori, debe ser estimado a partir de todos los posibles puntos de separación  $k$  que se encuentren maximizando  $G_{k,n}$  y se puede determinar a partir de la siguiente ecuación:

$$G_{max,n} = \max_k G_{k,n} \quad (2.80)$$



Por último, Hawkins y Zamba (2005b) proponen una adaptación a esta formulación para utilizarla en el control estadístico de procesos (SPC), donde el tamaño de la muestra no es fijo, sino que crece indefinidamente, es decir, crearon el procedimiento en curso de SPC como sigue:

- Calcular  $G_{max,n}$  después de que la observación  $n$  se ha añadido al registro total del proceso.
- Si  $G_{max,n} < h_n$ , donde  $h_n$  es un límite de control adecuado, entonces la conclusión es que no hay evidencia de un cambio en la media o varianza, y por lo tanto se debe dejar el proceso en ejecución ininterrumpida.
- Si  $G_{max,n} > h_n$ , se concluye que hay evidencia de un cambio en la media, la varianza, o ambos.

El hecho de usar el análisis de punto de cambio adaptado al procedimiento en curso de SPC, hace necesario la elección de los límites de control. Una opción atractiva para el enfoque del punto de cambio es tener una función de riesgo constante, mientras que el proceso se encuentra bajo control, es decir, se puede decir que se cumple lo siguiente



$$P\{G_{max,n} > h_{n,\alpha} | G_{max,j} \leq h_{j,\alpha}, j < n\} = \alpha \quad (2.81)$$

Donde  $\alpha$  es la probabilidad especificada de una señal errónea, y se denota el límite de control como  $h_{n,\alpha}$  para destacar esta dependencia. Debido a que esta probabilidad es constante, ésta corresponde a una ARL bajo control de  $1/\alpha$ . Por otra parte, no parece posible resolver estos valores  $h_{n,\alpha}$  teóricamente, por lo que Hawkins y Zamba (2005b) los estimaron utilizando la simulación. Éstos, utilizaron 10 millones de muestras aleatorias de la serie de datos  $N(0,1)$  de longitud de hasta 500, y que abarca los valores de  $\alpha$  0.05, 0.02, 0.01, 0.005, 0.002 y 0.001. El resultado de los fractiles estimados tiene errores estándar de aproximadamente 0.02.

Debido a que debe haber un mínimo de dos observaciones en cada segmento, es lógicamente posible empezar las pruebas desde la cuarta observación, pero esto es poco probable en la mayoría de las aplicaciones.

En el apéndice A, Tabla A.7 se muestran una parte de las simulaciones hechas por Hawkins y Zamba (2005b), en una tabla para obtener los límites de Control  $h_{n,\alpha}$  para una muestra de tamaño  $n$ , rechazo  $\alpha$ , a partir de un número de observaciones de  $n = 10$ .





Para facilitar el procedimiento, Hawkins y Zamba (2005b) sugieren el siguiente enfoque para la obtención de límites de control:

- Comparar las cinco primeras pruebas estadísticas de  $G_{max,n}$  (es decir, para  $n = 10, \dots, 14$ ) con las cinco primeras entradas de la tabla A.7. del apéndice.
- Luego, para  $n < 14$  se utiliza la siguiente aproximación, que reproduce la tabla de los valores estimados por Hawkins y Zamba para  $h_{n,\alpha}$ , con una desviación máxima absoluta de 0.09:

$$h \approx \begin{cases} 1.58 - 2.52 \log(\alpha) + \frac{0.094 + 0.33 \log(\alpha)}{\sqrt{n-9}} & \text{si } 0.001 \leq \alpha < 0.05 \\ 8.43 + 0.074 \log(n-9) & \text{si } \alpha = 0.05 \end{cases} \quad (2.82)$$

Ahora, si el gráfico emite una señal, es necesario hacer una investigación a fondo para conocer qué fue lo que exactamente ha cambiado entre los dos segmentos. En primer lugar, se podría comenzar con el fraccionamiento de la historia del proceso en el punto de cambio estimado, utilizando pruebas paramétricas o no paramétricas para llevar a cabo una comparación de dos muestras entre los dos segmentos resultantes. Se puede comenzar tomando el modelo de distribución normal del punto de cambio en



su valor nominal y la prueba de un cambio en la media y un cambio en la varianza utilizando los métodos convencionales de distribución normal.

Si el punto de cambio  $k$  ha sido detectado, Hawkins y Zamba (2005b) sugieren la aplicación de las siguientes pruebas paramétricas convencionales:

- Las dos caras de la prueba  $F$  para probar un cambio de varianza utilizando los grados de libertad  $k - 1$  y  $n - k - 1$  y el estadístico de prueba  $F = V_{0,k}(n - k - 1) / ((k - 1)V_{k,n})$ .

- La prueba  $t$  aproximada para probar un cambio en la media usando el estadístico  $t$  aproximado:

$$t = \frac{\bar{X}_{0,k} - \bar{X}_{k,n}}{\sqrt{S_{0,k} / (k - 1) + S_{k,n} / (n - k - 1)}} \quad (2.83)$$

La cual sigue aproximadamente una distribución  $t$  con  $r$  grados de libertad, donde:



$$r = \frac{\frac{S_{0,k}}{k-1} + \frac{S_{k,n}}{n-k-1}}{\frac{1}{k-1} + \frac{1}{n-k-1}} \quad (2.84)$$

El rendimiento de los gráficos de control de calidad se suele medir por su ARL. Si la función de riesgo fuera constante, entonces la ARL puede completamente caracterizar la distribución de la longitud de corrida. Hawkins (2005b) establece que longitud de corrida del enfoque del APC no es geométrico, ni tampoco parecen tener el peso de la cola lejos de la línea con los de la distribución geométrica. Por tanto, se puede comparar el rendimiento del método del análisis de punto de cambio con un enfoque de CUSUM.

Mediante simulaciones para un ARL de 500 bajo control. Se obtuvo los resultados de la Tabla 2.5. Todos tienen errores estándar de aproximadamente el 1% de su valor. El valor  $\delta = 0$ ,  $\sigma = 1$  representa el establecimiento bajo control, por lo que la ARL debería ser de 500, ya que están dentro de los errores de muestreo aleatorio.

La investigación en la columna  $\sigma = 1$ , pone de manifiesto la capacidad del procedimiento para detectar cambios en la media solamente. Si el proceso tiene corrida bajo control de 250 períodos, entonces la detección de incluso cambios modestos, es bastante rápido, pero en períodos de control cortos, junto con los cambios modestos son mucho más lentos de detectar. Las filas



con  $\delta = 0$  demuestran la capacidad del procedimiento para detectar cambios de variación pura. Al igual que con cambios de pura media, hay poder bastante bajo si el período inicial en el control es muy corto, pero mucho más alto si es incluso 50.

La capacidad modesta para detectar cambios pequeños que ocurren poco después del inicio del monitoreo es una fuerza de la formulación del punto de cambio con parámetros desconocidos, la que puede controlar su ARL durante el seguimiento de corridas cortas. Sería demasiado esperar a que lo haga con la misma sensibilidad que los métodos que requieren tiempo, cuidadosamente controlados en la fase I de los estudios.

Alejándose de la “ $T$ ” de la tabla, primera fila con cuarta columna, se muestra el desempeño del procedimiento para detectar cambios en la media y la varianza, que es generalmente mejor que el desempeño en la detección de cualquier partida sola. Dada que las particularidades del estadístico de la GLR combina las particularidades por separado de sus componentes  $t$  y  $F$ .

**Tabla 2.5.** ARL bajo control y fuera de control en el APC.

		ARL						
		$\sigma$						
$\delta$	$\tau+1$	0.51	0.64	0.8	1	1.25	1.56	1.95
0	10	287.6	415.9	480.4	496.6	508.7	500	471.9
	50	32.3	123.2	393.1	498.4	458.1	205.4	30.5



	<b>250</b>	23.2	44.6	186.6	491.1	204	34.2	14.2
<b>0.5</b>	<b>10</b>	180	321.1	422.3	743.7	492.7	477.7	448.1
	<b>50</b>	20.2	39.3	123.1	265.1	250.4	105.4	22.8
	<b>250</b>	16.8	25.2	41.4	63.7	47.6	23.3	12.4
<b>1</b>	<b>10</b>	47.2	120.5	228.3	355	410.6	419	407.2
	<b>50</b>	11	13.9	18.8	25	28.6	22.1	13.6
	<b>250</b>	10	12	14.3	16.2	15.7	12.8	9.2
<b>1.5</b>	<b>10</b>	12.3	21	55.2	132.7	230.9	298.6	316.9
	<b>50</b>	7	8	9	10.1	10.7	10.4	8.8
	<b>250</b>	6.5	7.1	7.8	8.2	8.2	7.7	6.7
<b>2</b>	<b>10</b>	7.2	8.8	13.6	26.9	65.8	132.2	178.9
	<b>50</b>	5	5.4	5.8	6.3	6.6	6.5	6.1
	<b>250</b>	4.6	4.9	5.2	5.4	5.4	5.3	5

Fuente: Hawkins (2005b).

El segundo problema de rendimiento se refiere al seguimiento de los diagnósticos realizados después de una señal para decidir si esto se debía a un cambio en la media, un cambio en la varianza, o un cambio en la media y la varianza. La tabla 2.4 sigue el esquema de la tabla 2.5 y muestra el porcentaje de veces que el seguimiento de la aproximación de la prueba  $t$  para la igualdad de la media alcanzada en el nivel de significación nominal de 1% y el porcentaje de veces que la prueba  $F$  alcanzó 1% de significancia



nominal. Las tres entradas para  $\delta = 0$ ,  $\sigma = 1$  muestran que en el caso nulo, el seguimiento de los diagnósticos de un cambio en la media de los casos  $80\%$  y un cambio de varianza en torno al  $90\%$  de los casos.

Los patrones de la tabla son bastante complejos. Disminuyendo en la columna  $\sigma = 1$ , la proporción de las señales de la varianza disminuyen, como era de esperar. La proporción de las señales de la media aumenta, pero luego disminuye. Esta disminución se produce porque la corrida corta antes de su descubrimiento, debido a combinación de pruebas que dan una señal antes de que las observaciones suficientes se hayan acumulado para el componente de  $t$  para alcanzar significación estadística.

En cuanto a la fila  $\delta = 0$  muestra que, la afirmación de una disminución de la varianza es más fiable, que la de un aumento de la varianza del mismo tamaño, aunque todos muestran una tasa de señal de al menos el  $85\%$ .

El paso a la porción “no- $T$ ” de la tabla muestra que la aparición simultánea de un cambio en la media reduce la capacidad de diagnosticar un cambio en la varianza. Un aumento en la varianza perjudica la capacidad de diagnosticar en la media, pero una reducción de la varianza lo mejora. Ésta indeterminación en la causa de la señal proviene de la longitud de las corridas cortas a la señal en estos entornos. Aunque podría parecer un



inconveniente que se puede tener una señal sin un diagnóstico claro, es más bien un homenaje a la forma en que el método combinado es capaz de detectar un problema antes de que sus componentes de pruebas individuales lo hagan.

#### 2.1.10. ESQUEMA DEL GRÁFICO DE CONTROL EWMA PARA OBSERVACIONES AUTOCORRELACIONADAS.

Según Lu y Reynolds (1999), cuando los datos se encuentran autocorrelacionados se aproximan a un modelo ARMA (1,1) o a un modelo AR(1).

Reynolds, Arnold y Baik (1996), hicieron ecuaciones para expresar los parámetros  $\phi, \theta, \sigma_{\gamma}^2$  en un modelo ARMA (1,1) en términos de los parámetros  $\phi, \sigma_{\gamma}^2, \sigma_{\varepsilon}^2$  en el modelo de errores aleatorios AR (1) y viceversa.

En particular si  $\phi > 0$  y  $\sigma_{\varepsilon}^2 > 0$ , entonces el modelo ARMA (1,1) con parámetros  $\theta$  y  $\sigma_{\gamma}^2$  puede obtenerse para un modelo de errores aleatorios AR(1) usando los parámetros:



$$\theta = \frac{\sigma_{\alpha}^2 + (1 + \varphi^2)\sigma_{\varepsilon}^2}{2\varphi\sigma_{\varepsilon}^2} - \frac{1}{2} \sqrt{\frac{\sigma_{\alpha}^2 + (1 + \varphi^2)\sigma_{\varepsilon}^2}{\varphi\sigma_{\varepsilon}^2}} \quad (2.85)$$

y

$$\sigma_{\gamma}^2 = \frac{\varphi\sigma_{\varepsilon}^2}{\theta} \quad (2.86)$$

Alternando si  $\varphi > 0$ , en un proceso AR(1), los parámetros se pueden obtener con un modelo ARMA (1,1) usando:

$$\sigma_{\gamma}^2 = \frac{\sigma_{\gamma}^2(\varphi - \theta)(1 - \varphi\theta)}{\varphi} \quad (2.87)$$

$$\sigma_{\varepsilon}^2 = \frac{\theta\sigma_{\gamma}^2}{\varphi} \quad (2.88)$$

Si los datos se ajustan mejor a un modelo AR(1), se estiman los parámetros de este modelo los cuales son:  $\varphi, \sigma_{\alpha}^2, \sigma_{\varepsilon}^2$ , una vez obtenido los parámetros se calcula  $\hat{\sigma}_{\mu}^2 = \frac{\sigma_{\alpha}^2}{(1 - \varphi^2)}$  para luego obtener  $\sigma_x^2 = \sigma_{\mu}^2 + \sigma_{\varepsilon}^2$ .

Si se aproxima los datos a un modelo ARMA(1,1) se estiman los parámetros del modelo, estos parámetros son:  $\varphi, \theta, \sigma_{\gamma}^2$ , una vez obtenido





estos parámetros se calcula  $\hat{\sigma}_\mu^2 = \sigma_\alpha^2 / (1 - \phi^2)$  y luego obtener

$$\sigma_x^2 = \sigma_\mu^2 + \sigma_\varepsilon^2.$$

### 2.1.10.1. LÍMITES DE CONTROL DEL GRÁFICO EWMA PARA DATOS AUTOCORRELACIONADOS

Los valores a graficar se obtendrán de la siguiente fórmula:

$$Y_t = \lambda X_t + (1 - \lambda) Y_{t-1}$$

(2.89)

Donde:

$X_t$  = valor observado individualmente o media muestral ( $\bar{X}_t$ )

$\lambda$  = es una constante de suavización.

Los límites de control serán los siguientes:

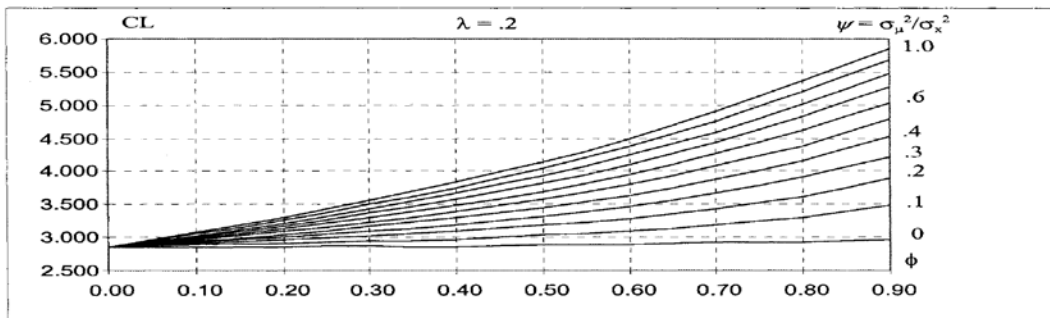
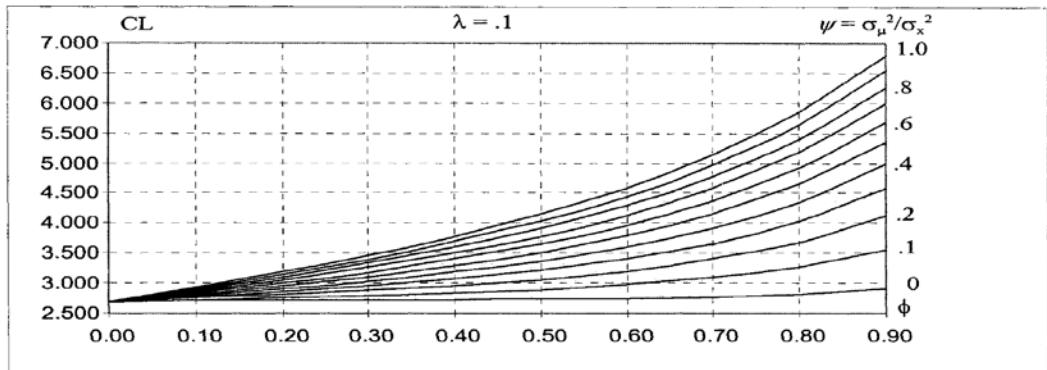
$$LIM = \mu_0 \pm c \sigma_x \sqrt{\frac{\lambda}{2 - \lambda}}$$

(2.90)

Los valores de  $\lambda$  y  $c$  se obtienen en función de los parámetros obtenidos de los modelos de series de tiempo. Los valores de  $c$  se pueden obtener a partir de las gráficas que se muestran en la figura 2.3, donde hay



diferentes valores de  $c$  para distintos  $\lambda, \varphi$  y  $\Psi$ . Puede ver la tabla A.3, que tiene diferentes valores de  $c$  extraídos de las gráficas antes mencionadas y la tabla A.4, que tiene diferentes valores de  $\lambda$  para diferentes valores de  $\varphi$  y  $\Psi$ . Ambas tablas las podrá encontrar en el apéndice I.



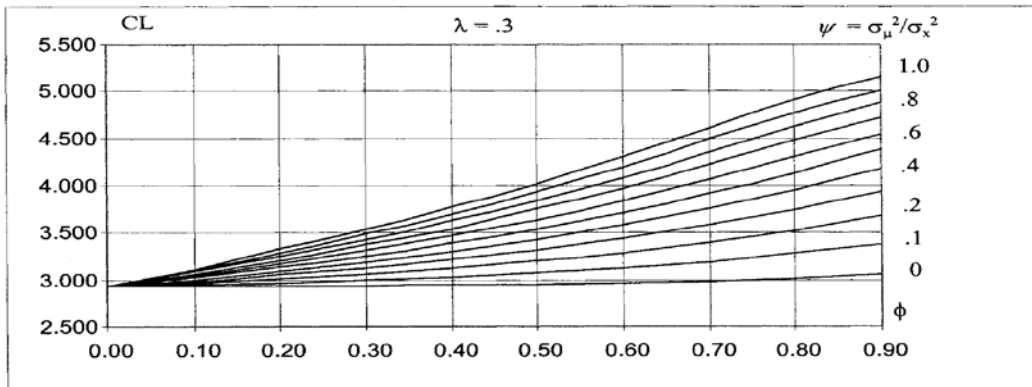


Figura 2.3. Valores de  $c$  a partir de diferentes valores de  $\lambda$ ,  $\phi$  y  $\Psi$ .

Fuente: Chao-Wen Lu y Mation R. Reynolds (1999).

### 2.1.11. ESQUEMA DEL GRÁFICO DE CONTROL CUSUM PARA OBSERVACIONES AUTOCORRELACIONADAS

Según Lu y Reynolds (2001), los gráficos de control para monitorear procesos son designados y evaluados tradicionalmente bajo la asunción de que las observaciones del proceso en diferentes tiempos son independientes. Sin embargo, la autocorrelación puede presentarse en muchos procesos, y puede impactar sobre las propiedades de dichos gráficos de control. Ellos plantean entonces un gráfico de control para monitorear la media del proceso para la situación en la que las observaciones son autocorrelacionadas y pueden ser modeladas como un modelo AR(1) con un error adicional.

Un gráfico CUSUM para detectar un cambio en un parámetro del proceso usualmente es definido como un gráfico de una cara. Un gráfico de dos caras es obtenido normalmente usando dos gráficos de una cara simultáneamente.



Primero se considera el gráfico de una cara basado en las observaciones originales. El estadístico para detectar incrementos en la media está definido por:

$$CO_k^+ = \max\{0, (CO_{k-1}^+ + (X_k - \mu_0 + r\sigma_x))\} \quad (2.91)$$

Donde:

$r^3 \theta$  = es una constante.

$r\sigma_x$  = es usualmente llamado el valor de referencia. El valor inicial de  $CO_k^+$  puede ser tomado como  $\theta$  (cero). Una señal es dada si  $CO_k^+$  cae por encima del límite de control,  $c\sigma_x$  cuando  $c$  es una constante.

El estadístico de una cara para detectar un decrecimiento en la media, se llama  $CO_k^-$  y viene definido por:

$$CO_k^- = \min\{0, (CO_{k-1}^- + (X_k - \mu_0 + r\sigma_x))\} \quad (2.92)$$

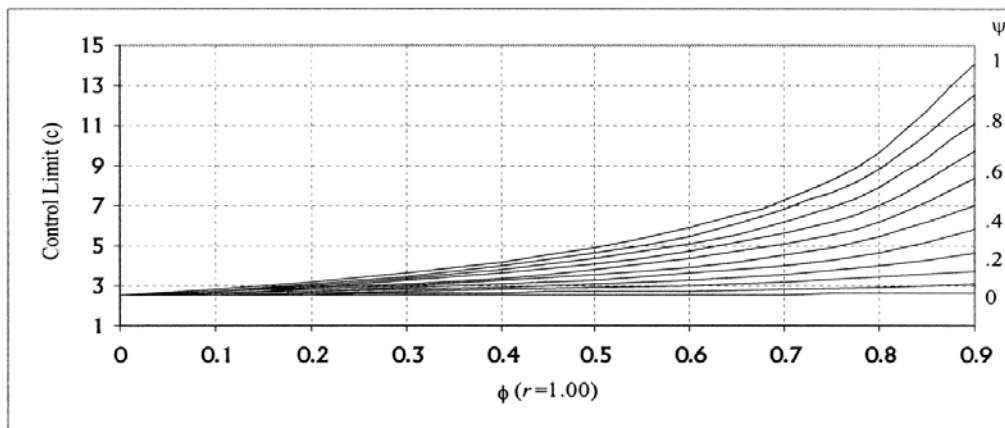
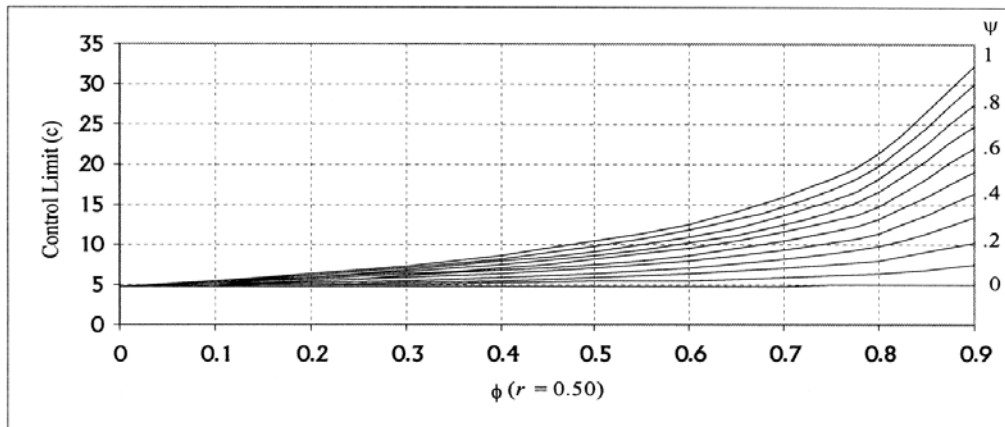
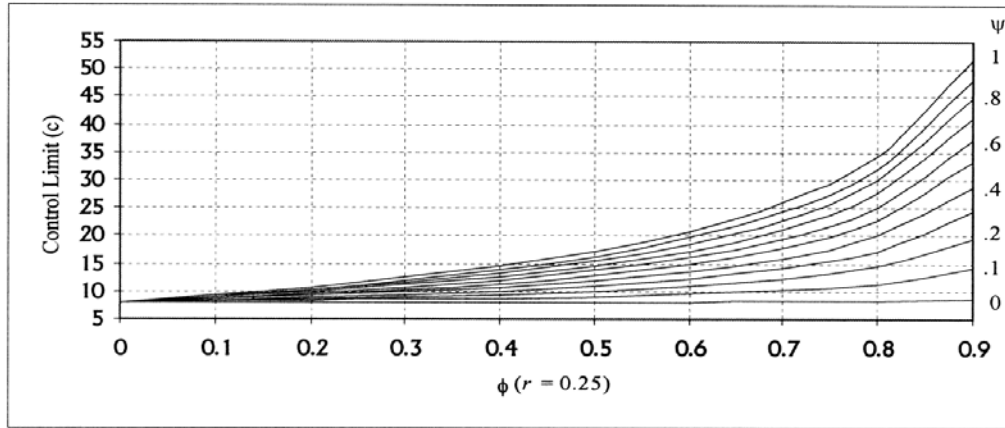
Donde  $r\sigma_x$  es el valor de referencia. Una señal es dada si  $CO_k^-$  cae por debajo del límite inferior de control  $-c\sigma_x$  donde  $c$  es una constante.



El gráfico CUSUM de dos caras usa a  $CO_k^+$  y  $CO_k^-$  simultáneamente. El gráfico  $CO_k^+$  mostrará valores por encima de cero y el gráfico  $CO_k^-$  mostrará valores por debajo de cero. Para diseñar un gráfico CUSUM en particular se deben obtener los valores de  $r$  y  $c$ .

La figura 2.4 proporciona valores de  $c$  que darán un ARL bajo control de 370 para varios valores de  $r, \varphi, \psi$  para observaciones de procesos que siguen el modelo AR(1).

Una vez que  $\varphi$  y  $\psi$  son estimados para una aplicación en particular, y el valor de  $r$  es obtenido, la figura 2.3 puede ser usada para determinar un apropiado valor de  $c$ . Vea la tabla A.5 del apéndice I que tiene diferentes valores de  $c$  obtenidos de la figura 2.4 y vea la tabla A.6 del apéndice I que muestra los diferentes valores de  $r$  dado  $\varphi$  y  $\psi$  y un cambio que se desea detectar más rápidamente de la media.





**Figura 2.4.** Valores de  $c$  para diferentes valores de  $r, \varphi, \psi$ .

**Fuente:** Chao-Wen Lu y Mation R. Reynolds (1999).

### 2.1.12. COMPORTAMIENTO DE LOS DATOS EN LOS PROCESOS PRODUCTIVOS.

Al referirse al comportamiento de los datos, Ferrer (2004) sostiene lo siguiente:

La aplicación del Control Estadístico de Procesos (Statistical Process Control, SPC) en procesos con dinámica, típicos en los modernos entornos altamente automatizados de la industria de piezas, se presenta donde la correlación entre las observaciones forma parte del sistema de causas comunes de variabilidad. Seguidamente se introduce la estrategia combinada ESPC (Engineering Statistical Process Control), integración del SPC con la teoría de control de procesos (Engineering Process Control, EPC), que aprovecha la información valiosa que proporciona la propia dinámica de los procesos, junto con la existencia de variables de control, fácilmente manipulables, para mejorar el control de los procesos.

Entre los métodos del SPC, destacan los gráficos de control; en el empleo y diseño de estos gráficos se asume que las observaciones son estadísticamente independientes, siendo ésta una hipótesis clave en las



propiedades estadísticas de estos gráficos. Sin embargo, esta suposición es frecuentemente violada en muchos procesos industriales modernos. Así, aunque en la industria de piezas (automóvil, electrónica, etc.) tradicionalmente ésta hipótesis ha sido razonable en la práctica, por lo que los gráficos de control han resultado excelentes herramientas para mantener los procesos bajo control estadístico, sin embargo, con el desarrollo de las nuevas tecnologías que facilitan la automatización de los procesos, es frecuente encontrar dispositivos que proporcionan en un entorno CIM (Computer Integrated Manufacturing) información sobre cada pieza fabricada, o sobre parámetros operativos del proceso (v.g. presión del circuito hidráulico del brazo del robot que debe taladrar la pieza), no pudiendo aceptarse como válida la hipótesis de independencia en los datos. Esto mismo ocurre, por otra parte, en las industrias de procesos con procedimientos de fabricación en continuo o por lotes, típicos de sectores como el petroquímico o el farmacéutico, donde es frecuente la existencia de autocorrelaciones entre medidas sucesivas.

La presencia de correlación no debe interpretarse como una causa especial de variabilidad, sino como parte del sistema de causas comunes intrínseco al proceso en cuestión, consecuencia de la presencia de elementos inerciales como flujo de materias primas, tanques de almacenamiento, reactores, reflujos, condiciones ambientales, etc. con tiempos de residencia o dinámicas (inercias) mayores que la frecuencia de muestreo. Por ejemplo, si en un proceso de fabricación de polietileno mediante un reactor en lazo, las





moléculas de etileno tienen un tiempo de residencia en el reactor de unas cuatro horas y se toma una muestra del reactor cada hora, hay moléculas que han coexistido en el reactor en varias tomas de muestra, por lo que es esperable encontrar correlaciones entre muestras consecutivas (Capilla et al. 1999). En general todos los procedimientos de fabricación presentan cierta inercia frente a la acción de parámetros ambientales, el cambio de materias primas o al ajuste de sus condiciones operativas, por lo que un aumento de la frecuencia de muestreo puede llegar a generar la aparición de autocorrelación entre observaciones cercanas en el tiempo. (p. 20-21)

### **2.1.13. LENGUAJE DE PROGRAMACIÓN MATLAB.**

MATLAB es el nombre abreviado de “MATrix LABoratory” y es un programa diseñado especialmente para realizar cálculos numéricos con vectores y matrices. MATLAB tiene también un lenguaje de programación propio. Puede trabajar con números escalares, tanto reales como complejos, con cadenas de caracteres y con otras estructuras de información más complejas. Una de las capacidades más atractivas es la de realizar una amplia variedad de gráficos en dos y tres dimensiones.

MATLAB es un gran programa de cálculo técnico y científico y siempre es una magnífica herramienta de alto nivel para desarrollar aplicaciones técnicas, fácil de utilizar y que aumenta significativamente la productividad de



los programadores respecto a otros entornos de desarrollo debido a que dispone de un código básico y de varias librerías especializadas denominadas “toolboxes”.

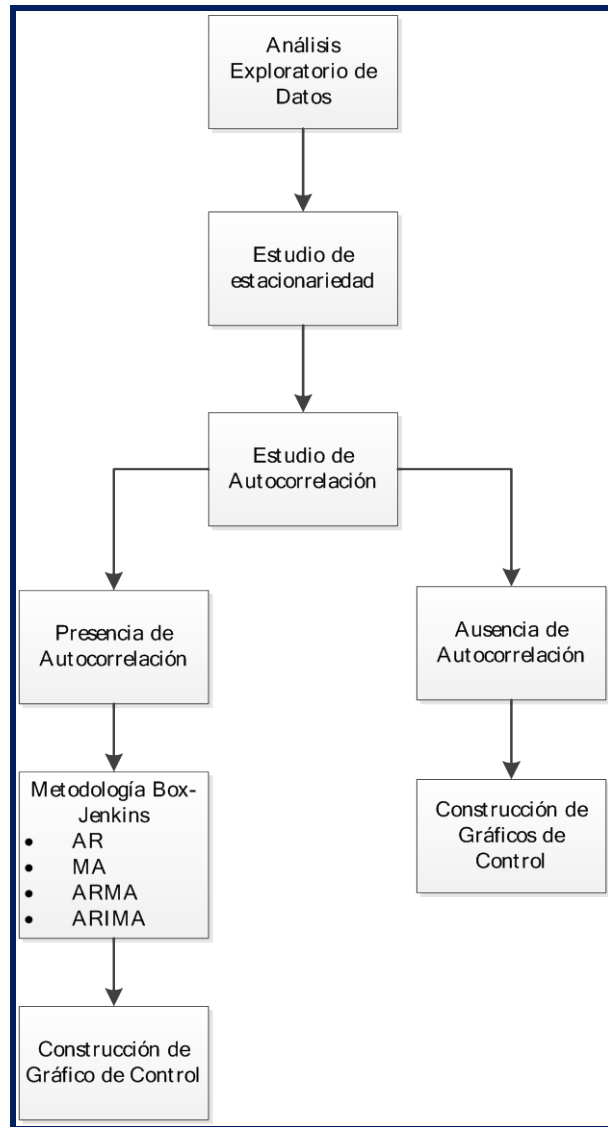


### **3. METODOLOGÍA.**

El presente capítulo muestra una metodología que indica los pasos a seguir en el análisis de datos exploratorio, basados en diversas pruebas estadísticas que pueden ser aplicadas para determinar qué tipo de comportamiento presenta y la posterior construcción de gráficos de control. La metodología propuesta se indica en la figura 3.1.

#### **3.1. METODOLOGÍA PROPUESTA.**

La metodología propuesta en el presente trabajo de grado se realiza en cuatro fases: análisis exploratorio de datos, estudio de estacionariedad, estudio de autocorrelación y por último la construcción de gráficos de control. A continuación se detallan los procedimientos utilizados.



**Figura 3.1.** Procedimiento seguido en ésta investigación para identificar el modelo univariante.

**Fuente:** Elaboración Propia.



### **3.1.1. ANÁLISIS EXPLORATORIO DE LOS DATOS.**

#### **3.1.1.1. PRUEBA DE ALEATORIEDAD.**

Para comprobar que un proceso sea aleatorio, el primer paso es realizar una verificación de independencia de los datos. Existen diversas pruebas para comprobar ésta independencia como se puede ver en el trabajo de Cromwell, Labys, Terraza (1994); en esta investigación se realiza el estudio mediante el contraste de rachas, por estar incluida en el paquete estadístico de MATLAB, lo que permite verificar la hipótesis nula de si las sucesivas observaciones son independientes.

#### **3.1.1.2. ANÁLISIS GRÁFICO.**

Luego, con el fin de tener una idea inicial de la información contenida en el conjunto de datos se realiza el análisis estadístico gráfico para observar el comportamiento, tendencia, distribución y la posible presencia de datos atípicos de los datos. Para tal fin, se puede apoyar el análisis en las herramientas gráficas como el histograma de frecuencia, el cual indica que la superficie de cada barra es proporcional a la frecuencia de los valores representados. Otra de las herramientas utilizadas en el análisis grafico es el



diagrama de caja, basado en cuartiles, que proporciona una visión general de la simetría de la distribución de los datos y la presencia de valores atípicos. En conjunto con el gráfico de probabilidad normal, la cual es utilizada para contrastar la normalidad de un conjunto de datos, se puede tener una buena base de apoyo visual en el análisis de los datos.

### **3.1.1.3. DETECCIÓN Y ELIMINACIÓN DE OUTLIERS.**

El siguiente paso es la detección de los valores atípicos. A pesar de la existencia de diversas pruebas para la detección de estos valores, en este caso se utilizó la prueba de Rosner (1975) debido a que tiene la capacidad de detectar múltiples outliers.

Varios autores, entre ellos Hawkins (1980), recomiendan dicha prueba en contraste a otras pruebas de detección de outliers ya que la primera detecta  $k$  valores atípicos a diferencia de otras.

### **3.1.1.4. PRUEBA DE NORMALIDAD.**

Como paso final en la etapa de análisis exploratorio de la muestra, se procede a realizar pruebas de normalidad; debido a que diversas pruebas



necesarias para la construcción de los gráficos de control aquí tratados fueron formuladas bajo el supuesto de normalidad. La prueba estadística de **Lilliefors** es la herramienta escogida para éste estudio, no solo por su practicidad de ejecución sino que además viene incluido en la herramienta computacional MATLAB.



**Figura 3.2.** Diagrama de la fase de Análisis Exploratorio de los Datos.

**Fuente:** Elaboración propia.

### 3.1.2. ESTUDIO DE ESTACIONARIEDAD.

Para lograr el objetivo de diseñar los gráficos de control propuestos en ésta investigación, las series de tiempo deben ser datos estacionarios, es decir, no deben presentar cambios significativos en su media o la varianza. Por lo tanto, se debe garantizar que la data a utilizar es estacionaria.

Para ello se escogió la prueba de raíces unitarias de Dickey–Fuller ya que se puede observar en el proceso las variaciones o no, las diversas tendencias determinísticas o estocásticas.



### 3.1.2.1. ESTACIONARIEDAD EN LA VARIANZA.

El primer parámetro a analizar es la varianza, debido a que si no se controla este parámetro no se lograrían obtener resultados adecuados debido a la gran variabilidad que podría presentar.

Si la serie no es estacionaria se utilizará las transformaciones de Box-Cox para eliminar la no estacionariedad en la varianza.

### 3.1.2.2. ESTACIONARIEDAD EN LA MEDIA.

Una vez de asegurarse que la serie de tiempo es estacionaria en la varianza, el siguiente paso es el análisis de estacionariedad en la media.

Gujarati (2004), recomienda utilizar las siguientes pruebas estadísticas para detectar la estacionariedad en la media:

- La prueba de raíz unitaria Dickey - Fuller (DF).
- La prueba de raíz unitaria de Phillips - Perron (PP).

De ser necesario se hará la corrección de no estacionariedad en la media mediante transformaciones llamadas diferenciaciones, las cuales se





pueden realizar con un máximo de dos diferenciaciones para lograr una serie estacionaria en la media.

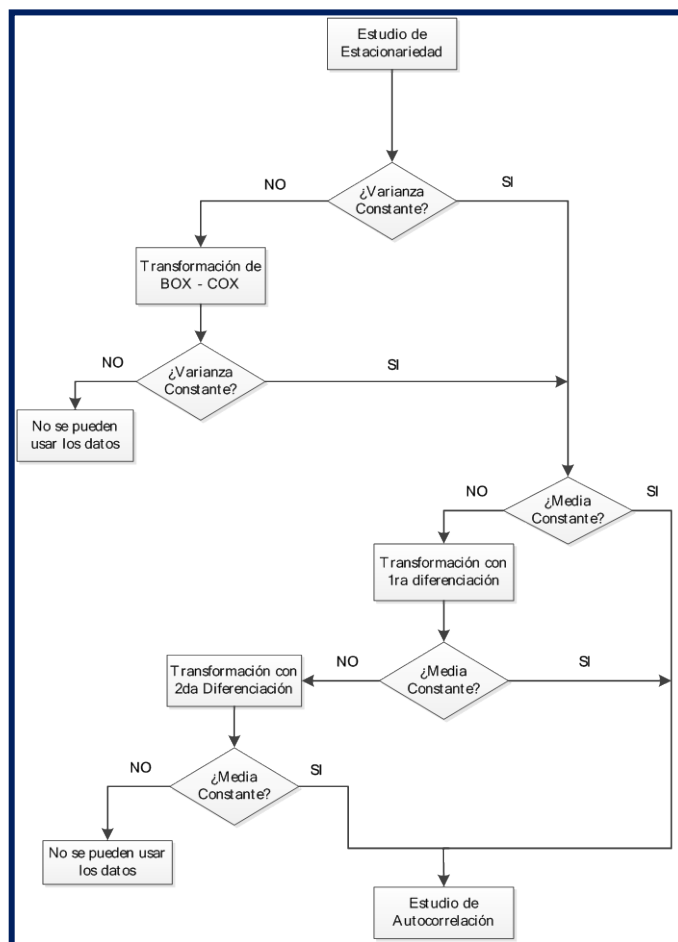


Figura 3.3. Diagrama de flujo del estudio de estacionariedad.

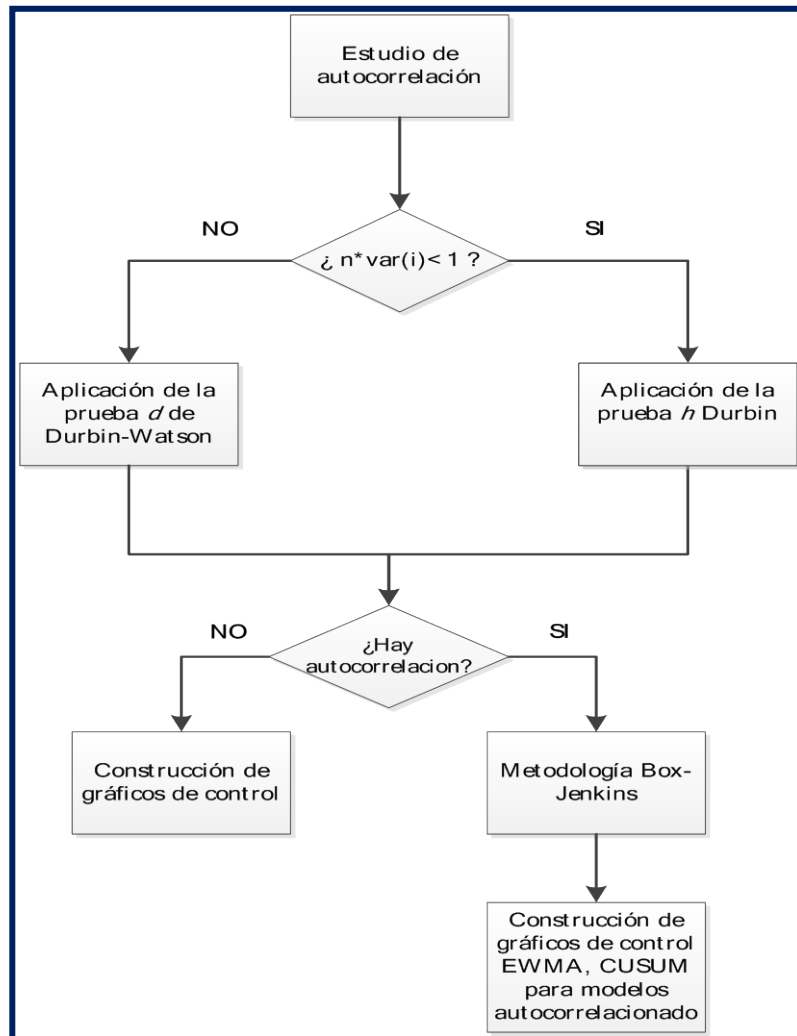
Fuente: Elaboración propia.



### 3.1.3. ESTUDIO DE AUTOCORRELACIÓN.

Para asegurar resultados confiables en la escogencia del modelo de la serie de tiempo a utilizar, la determinación de la presencia o no de autocorrelación es una fase de importancia. Para este estudio se utiliza primeramente el contraste  $h$  de Durbin, por ser una prueba de mayor confiabilidad, pero antes se tiene que verificar cierta condición para su aplicación. De no cumplirse la condición se procederá a utilizar la prueba de Durbin-Watson; la cual, es una prueba que viene dada en el paquete estadístico MATLAB.

Luego de aplicado el contraste conveniente, si la serie no presenta autocorrelación, se asume que los datos son independientes y por tanto pueden ser usados en el diseño de los gráficos de control y Análisis de punto de cambio. En caso contrario, los datos pasan a efectuarles un análisis de serie de tiempo por el método de Box y Jenkins.





**Figura 3.4.** Diagrama de flujo del estudio de autocorrelación.

**Fuente:** Elaboración propia.

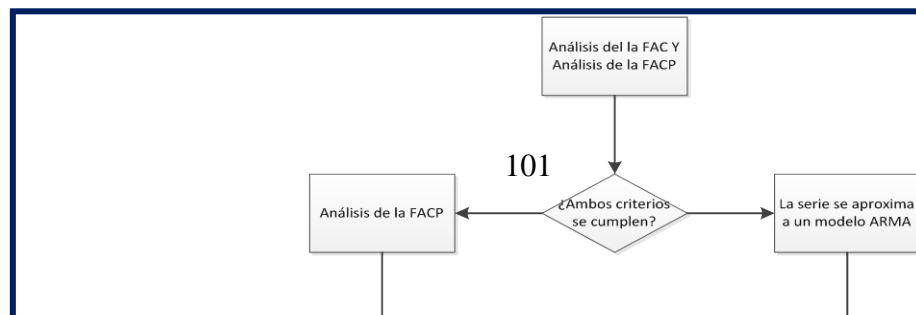
### **3.1.4. ELABORACIÓN DEL MODELO DE SERIE DE TIEMPO APROPIADO PARA DATOS AUTOCORRELACIONADOS**

El estudio consiste en graficar la función de autocorrelación y la función de autocorrelación parcial. Se identifican los procesos partiendo de las propiedades teóricas de los procesos AR o MA, según los cuales, un proceso  $AR(p)$  tiene una función de autocorrelación parcial que se anula a partir de un retardo  $p$ , y un proceso  $MA(q)$  tiene una función de autocorrelación que se anula a partir de un retardo  $q$ .

#### **3.1.4.1.1. REGLA DE DECISIÓN PARA ELECCIÓN DE MODELOS DE SERIES DE TIEMPO.**



- Si la Función de autocorrelación (FAC) presenta un decrecimiento rápido de tipo exponencial y / o sinusoidal la serie la aproxima a un modelo AR (p).
- Si la Función de autocorrelación Parcial (FACP) presenta un decrecimiento rápido de tipo exponencial y/o sinusoidal la serie la aproxima a un modelo MA (q).
- Si se cumple de forma conjunta las condiciones anteriores la serie se aproxima a un modelo ARMA (p, q).





**Figura 3.5.** Diagrama de flujo de la metodología de Box – Jenkins.

**Fuente:** Elaboración propia.

### 3.1.5. CONSTRUCCIÓN DE GRÁFICOS DE CONTROL.



	REQUISITOS PARA EL DISEÑO DE LOS GRÁFICOS DE CONTROL	COMPORTAMIENTO DE LOS DATOS EN PROCESOS PRODUCTIVOS
--	--	---

En la tabla 3.1. se realiza una comparación del posible comportamiento de los datos y los requisitos necesarios para hacer la construcción de los gráficos CUSUM, EWMA y APC.



<p><b>EWMA</b></p>	<p>Es robusto a la presencia de no normalidad en los datos.</p> <p>Es ajustado a cambios de una magnitud específica.</p> <p>Requiere por lo general de una costosa fase de recolección de muestras para estimar aproximadamente los parámetros bajo control.</p> <p>Gran cantidad de analistas se encuentran familiarizados con este tipo de gráficos.</p> <p>Trabaja bajo el supuesto de que los datos no presentan autocorrelación.</p>	<p>Los datos provenientes de gran parte de las industrias presentan :</p>
<p><b>CUSUM</b></p>	<p>Ofrece una estimación del punto donde aproximadamente ocurrió el cambio.</p> <p>Es sensible a la presencia de no normalidad en los datos.</p> <p>Es ajustado a cambios de una magnitud específica.</p> <p>Requiere por lo general de una costosa fase de recolección de muestras para estimar aproximadamente los parámetros bajo control.</p> <p>Gran cantidad de analistas se encuentran familiarizados con este tipo de gráficos.</p> <p>Trabaja bajo el supuesto de que los datos no presentan autocorrelación.</p>	<p>Tendencias estocásticas.</p> <p>Poca variabilidad en los datos.</p> <p>Presencia de autocorrelación.</p> <p>Variaciones estacionales (anual, semestral u otro).</p>
<p><b>APC</b></p>	<p>Ofrece una estimación del punto donde aproximadamente ocurrió el cambio.</p> <p>El estadístico GLR es sensible a la presencia de no normalidad en los datos.</p> <p>No requiere una gran fase de implementación.</p> <p>Es competitivo en la detección de causas especiales aisladas no sostenidas.</p> <p>Técnica para parámetros desconocidos.</p>	





	Trabaja bajo el supuesto de que los datos no presentan autocorrelación.	
--	---	--

**Tabla 4.1.** Comportamiento de los datos versus gráficos de control.

**Fuente:** Elaboración propia.

### 3.1.6. GRÁFICOS DE CONTROL EWMA PARA DATOS NO AUTOCORRELACIONADOS.

Para construir el gráfico de control EWMA, se necesita conocer de antemano los parámetros  $\lambda$  y  $k$ , con el fin de establecer los límites de control del gráfico, por tal motivo se debe comenzar realizando una estimación de ellos. Dichos parámetros se encuentran tabulados en tablas, sin embargo, para conseguirlos se necesita conocer el ARL y un cambio en la media ( $D$ ). Una vez que se tenga el cambio en la media ( $D$ ) y con el ARL correspondiente, se busca en la tabla correspondiente el valor de  $\lambda$  y  $k$ .

En la figura 3.6 se muestran los pasos a seguir en la construcción del gráfico de control EWMA.

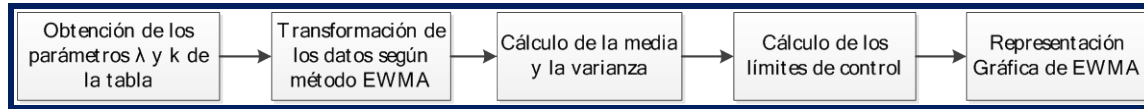
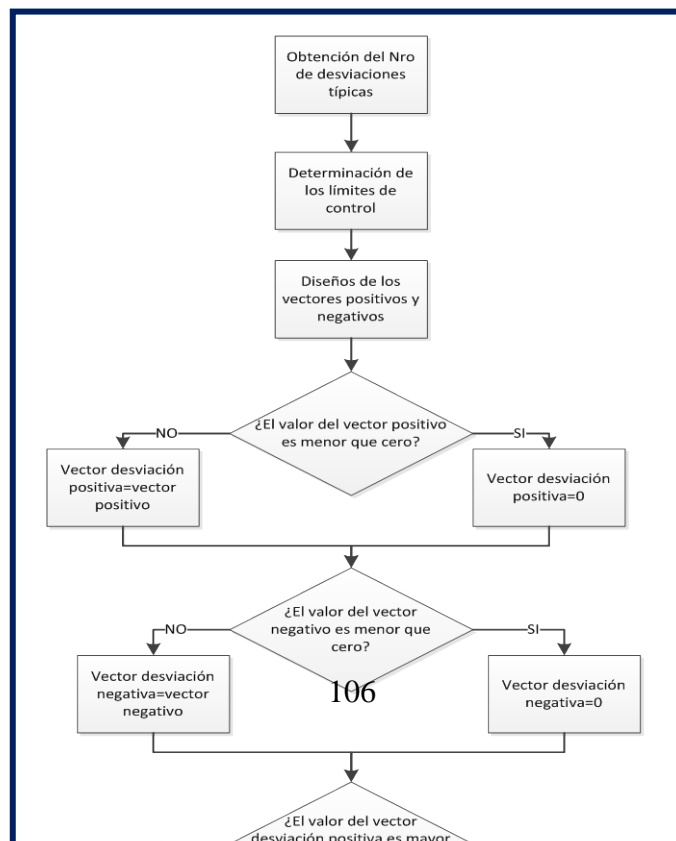


Figura 3.6. Diagrama de flujo del gráfico de control EWMA.

Fuente: Elaboración propia.

### 3.1.7. GRÁFICOS DE CONTROL CUSUM PARA DATOS NO AUTOCORRELACIONADOS.

Siguiendo la metodología para construir los límites de control del gráfico CUSUM, se tiene que establecer el número de desviaciones típicas que se desea para detectar cambios o corrimientos de esa magnitud. Tomando en cuenta que dicho número debe ser mayor que cero y generalmente es menor que dos debido a que estos gráficos presentan un buen rendimiento frente a corrimientos pequeños de la media.





**Figura 3.7.** Diagrama de flujo del gráfico de control CUSUM.

**Fuente:** Elaboración propia.

### **3.1.8. GRÁFICOS DE CONTROL EWMA PARA DATOS AUTOCORRELACIONADOS.**

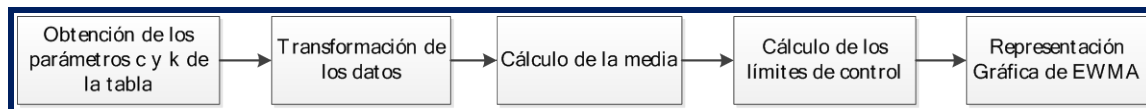
Habiendo obtenido los parámetros del modelo por medio de la metodología Box y Jenkins se necesita obtener los parámetros  $\lambda$  y  $c$  para



diseñar los límites de control y el vector  $Z$  (media móvil ponderado exponencialmente).

Estos valores están tabulados y se obtienen a partir de los valores de  $\Psi$ , de  $\phi$ , que son los parámetros del modelo, y un cambio en la media ( $\delta$ ) que deberá elegir el analista arbitrariamente. Teniendo el valor de  $\lambda$  y los valores de  $\Psi$  y de  $\phi$ , el valor de  $c$  se obtiene a partir de otra tabla.

En la figura 3.8 se muestran los pasos a seguir en la construcción del gráfico de control.



**Figura 3.8.** Diagrama de flujo del gráfico de control EWMA autocorrelacionado.

**Fuente:** Elaboración propia.

### 3.1.9. GRÁFICOS DE CONTROL CUSUM PARA DATOS AUTOCORRELACIONADOS.

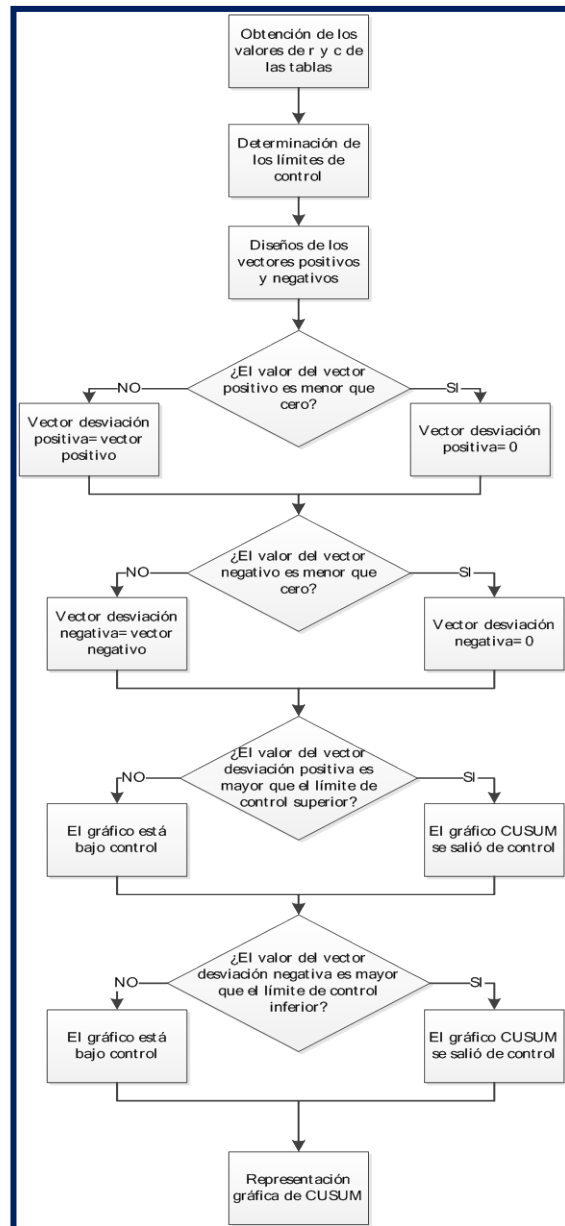
Lo primero que se realiza es obtener los valores de  $r$  y  $c$  de las tablas correspondientes.



Estos valores están tabulados y se obtienen a partir de los valores de  $\Psi$ , de  $\phi$ , que son los parámetros del modelo, y un cambio en la media ( $\delta$ ) que deberá elegir el analista arbitrariamente y buscando el menor valor de ARL se consigue el valor de  $r$ .

Teniendo el valor de  $r$  y los valores de  $\Psi$  y de  $\phi$ , el valor se  $c$  se obtiene a partir de otra tabla.

Habiendo obtenido los valores de  $r$  y  $c$  el procedimiento que se lleva a cabo para obtener el gráfico es similar al CUSUM para datos independientes. El único cambio es que ahora no se tendrá que establecer el número de desviaciones típicas, porque este valor viene representado por el valor de  $r$  ya determinado.





**Figura 3.9.** Diagrama de flujo del gráfico de control CUSUM autocorrelacionado.

**Fuente:** Elaboración propia.

### 3.1.10. ANÁLISIS DE PUNTO DE CAMBIO.

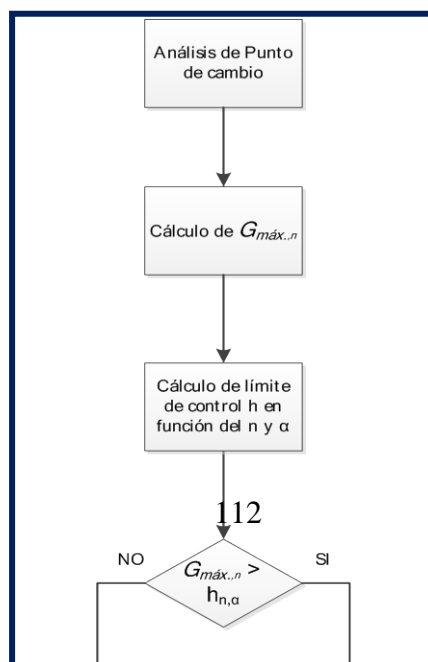
Hawkins y Zamba (2005) plantea una formulación para utilizarla en el control estadístico de procesos (SPC), donde el tamaño de la muestra no es fijo sino que crece indefinidamente, para ello usan límites de control ( $h_n$ ) en conjunto con la prueba estadística de Bartlett corregida ( $G_{k,n}$ ), la cual debe ser estimada a partir de todos los posibles puntos de separación  $k$  que se encuentren maximizando  $G_{k,n}$ , es decir, calculando  $G_{max,n}$ .

Para calcular el límite de control del análisis de punto de cambio, el analista debe indicar el valor de alfa ( $\alpha$  es la probabilidad especificada de



una señal errónea). Una vez seleccionado el valor de alfa, se obtienen los límites de control. Seguidamente se debe continuar con el cálculo del estimador  $G_{max,n}$ , para verificar si dicho valor sobrepasa el límite de control. En el caso de que lo haga, el analista debe identificar que fue lo que exactamente cambió utilizando las pruebas estadísticas para ello (Prueba  $t$  y Prueba  $F$  para dos muestras) para informar al analista si el cambio ocurrió en la media, en la varianza o en ambos.

A continuación se muestra la secuencia lógica que se debe llevar a cabo para aplicar el análisis de punto de cambio.







**Figura 3.10.** Diagrama de flujo del análisis de punto de cambio.

**Fuente:** Elaboración propia.



#### 4. DISEÑO DE LA HERRAMIENTA.

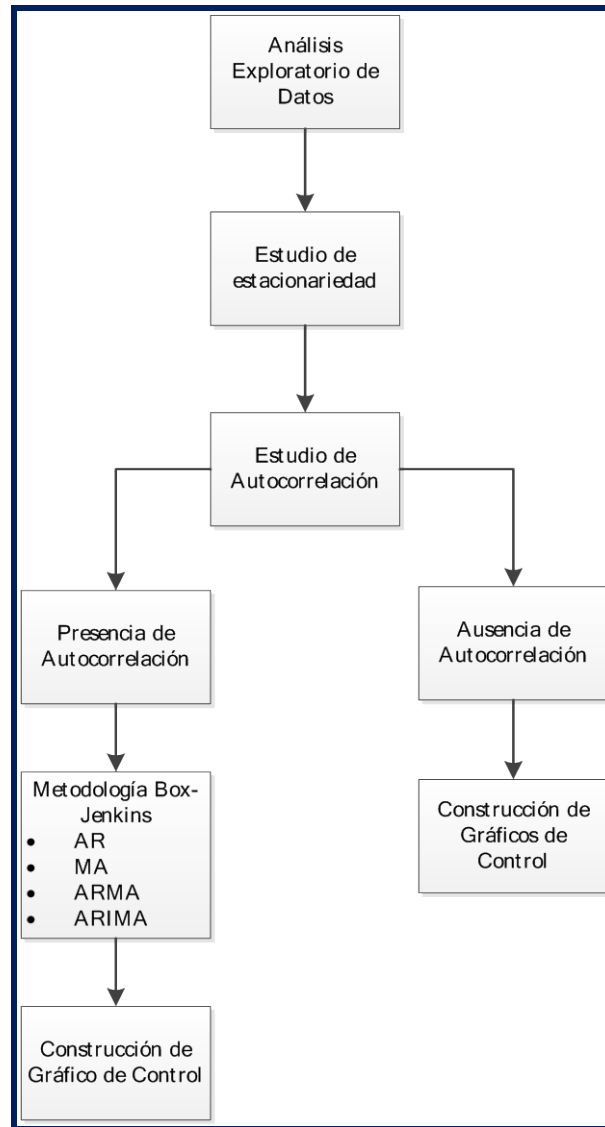
El presente capítulo expone de manera detallada la estructura de la herramienta computacional desarrollada en lenguaje MATLAB, con el fin lograr una mayor comprensión del programa y facilitar su uso.

La herramienta desarrollada, permite el análisis y depuración de los datos de tipo series de tiempo, la construcción de los gráficos de control CUSUM y EWMA para procesos estacionarios con ausencia de autocorrelación en forma conjunta con el Análisis de Punto de Cambio; además, de la construcción de los gráficos antes mencionados con evidencia de autocorrelación.

El programa principal y los subprogramas diseñados se encuentran en su totalidad en el Apéndice I.



A continuación se presenta la estructura general del programa:



**Figura 4.1.** Diagrama de flujo General.

**Fuente:** Elaboración propia.

Luego que los datos ingresen al sistema se ejecuta la etapa inicial del software, en donde se evalúa si los datos vienen dados como subgrupos y de ser así, los valores de cada muestra se promedian para obtener unos datos



de tipo vector para luego continuar con la siguiente fase correspondiente al análisis exploratorio de los datos, donde entre otros análisis se realiza el estudio para la detección y eliminación de posibles valores atípicos en la muestra. Luego se realiza un estudio de estacionariedad, donde, de ser necesario se aplicarán las transformaciones para que la serie sea estacionaria, posteriormente el análisis de autocorrelación permitirá conocer si los datos resultan ser autocorrelacionados. Si resultan ser autocorrelacionados, se realiza el estudio basado en la metodología de Box y Jenkins para determinar el modelo de serie de tiempo que más se aproxima a la data estudiada. Si los datos arrojan ser estacionarios y no autocorrelacionados se procede a diseñar los gráficos de control apropiados para tal tipo de muestra en conjunto con el análisis de punto de cambio.

#### **4.1. ENTRADA DE BASE DE DATOS.**

El software requiere que los nuevos datos posean una configuración específica, para que puedan ser leídos por el mismo, deben estar escritos en un archivo de tipo “.xls” para “EXCEL 1997–2003” o “.xlsx” para “EXCEL 2007–2010” y ordenados en subgrupos racionales sucesivos. Es importante el hecho que el archivo no debe tener caracteres alfabéticos, ni títulos en la hoja Excel, solo números.



Cada subgrupo deberá colocarse en las filas del archivo, y cada dato del subgrupo en las columnas.

#### **4.2. ANÁLISIS PRELIMINARES.**

Dado que los gráficos de control que serán empleados se valen de muestras univariantes, el tamaño del subgrupo racional debe ser igual a uno, en ésta etapa se verificará el tamaño del subgrupo racional, de modo tal que en caso de ser mayor a uno, deben ser transformados a través del cálculo de promedios de cada subgrupo para obtener datos tipo vector, es decir, se calcula el promedio de cada fila de la base de datos y dichos valores obtenidos formarán un vector columna que si puede ser introducido para el análisis.

Es importante el orden cronológico en la información. Se concluyó que generalmente no se conoce el orden cronológico de los datos por lo que se puede promediar el tamaño de la muestra y trabajar como un único tamaño.

#### **4.3. ANÁLISIS EXPLORATORIO DE LOS DATOS.**

##### **4.3.1. PRUEBA DE ALEATORIEDAD.**



Dentro de esta etapa se debe evaluar la aleatoriedad de los datos, para ello se hace uso de una función de MATLAB conocida como “runstest” ( $h = \text{runstest}(x)$ ), la cual realiza una prueba estadística en la secuencia de observaciones en un vector “ $x$ ” y se prueba la hipótesis nula de que los valores de “ $x$ ” vienen en orden aleatorio, contra la alternativa de que no lo hacen. La prueba se basa en el número de corridas de valores consecutivos por encima o por debajo de la media de “ $x$ ”. La prueba devuelve el valor lógico  $h = 1$  si se rechaza la hipótesis nula al nivel de significancia del 5%, y  $h = 0$  si no puede hacerlo. El programa desarrollado mostrará una advertencia en el caso de que exista evidencia suficiente de que los datos no son aleatorios (al rechazar la hipótesis nula), en caso contrario se emitirá un mensaje en donde se informa que la hipótesis nula se acepta y los datos efectivamente vienen de manera aleatoria.

#### 4.3.2. ANÁLISIS GRÁFICO.

Seguidamente, se procede a emitir en pantalla una serie de gráficos de los datos suministrados, con el fin que el usuario tenga una visualización en pantalla del comportamiento de la serie de tiempo sin ninguna transformación; para ello, el programa se apoya en varias funciones gráficas de MATLAB, donde en primer lugar se representa los valores de la muestra en el tiempo



con ayuda de la función conocida como “plot” ( $\text{plot}(x)$ ), luego continúa con la realización del histograma de frecuencia a partir de la función “histfit” ( $\text{histfit}(x)$ ), en la que ubican los elementos del vector  $x$  en el eje horizontal y la frecuencia de los mismos en el eje vertical. Inmediatamente se procede a realizar el diagrama de caja mediante la función “boxplot” ( $\text{boxplot}(x)$ ), donde si  $x$  es un vector, sólo hay una caja. En dicha caja, la marca central es la mediana, los bordes de la caja son los percentiles 25 y 75, los bigotes se extienden a los puntos de datos más extremos que no se consideran outliers y de haberlos se representan de forma individual; para completar el estudio gráfico se ejecuta el comando “probplot” ( $\text{probplot}(x)$ ) con el fin de graficar el contraste de normalidad del conjunto de datos. Permite comparar la distribución empírica de la muestra con la distribución normal.

#### 4.3.3. DETECCIÓN Y ELIMINACIÓN DE OUTLIERS.

El análisis del efecto que pueden llegar a tener los outliers en el comportamiento de la muestra es una fase de gran importancia en el análisis de datos. La presencia de estos pueden dar a entender un comportamiento distinto a si no están dentro de la muestra.

Varios autores, entre ellos Hawkins (1980), recomiendan la prueba de Rosner (1975) en contraste a otras pruebas de detección de outliers ya que





la primera detecta  $k$  valores atípicos a diferencia de otras. Esta prueba se basa en la hipótesis nula de que no hay presencia de outliers en la muestra contra la alternativa de que existen  $r$  outliers en la muestra.

En el programa este calculo esta representado por una función llamada “outliers” (`outliers(X,alpha,k)`), en el cual se lee el vector completo, se calcula el estadístico para la cantidad de los valores atípicos introducidos por el usuario. Luego, se acepta o rechaza la hipótesis planteada.

#### 4.3.4. PRUEBA DE NORMALIDAD.

Como último paso en la etapa de análisis exploratorio de la muestra, se procede a realizar la prueba de normalidad; debido a que diversas pruebas necesarias para la construcción de los gráficos de control aquí tratados fueron formulados bajo la suposición de normalidad, seguidamente se realiza un estudio de normalidad a la serie de tiempo para verificar el cumplimiento de esta asunción. En el estudio de normalidad, se le aplicará a la serie de tiempo en estudio, la prueba estadística de Lilliefors que ofrece MATLAB en la función “lillietest” (`lillietest(x)`) para probar la hipótesis nula de que la serie de tiempo en estudio proviene de una familia de la distribución Normal, contra la hipótesis alternativa de que no proviene de una distribución Normal, esta prueba se usara con un nivel de significancia de  $5\%$ .



#### **4.4. ESTUDIO DE ESTACIONARIEDAD.**

Para lograr el objetivo de diseñar los gráficos de control propuestos en ésta investigación, las series de tiempo deben ser datos estacionarios, es decir, que no deben presentar cambios significativos en su media o la varianza, ni componentes estacionales y/o con tendencia. Por lo tanto, se debe garantizar que la data a utilizar es estacionaria.

##### **4.4.1. ESTACIONARIEDAD EN LA VARIANZA.**

El primer parámetro a analizar es la varianza, debido a que si no se controla este parámetro no se lograrían obtener resultados adecuados debido a la gran variabilidad que podría presentar.

Se utilizará la familia de transformaciones de Box-Cox para eliminar la no estacionariedad en la varianza. Dicha transformación se realiza mediante una función llamada “boxcox”(  $\text{boxcox}(X)$  ) que esta contenida en Matlab.

##### **4.4.2. ESTACIONARIEDAD EN LA MEDIA.**

Una vez de asegurarse que la serie de tiempo es estacionaria en la varianza, el siguiente paso es el análisis de estacionariedad en la media.



Según Gujarati (2004), recomienda utilizar las siguientes pruebas estadísticas para detectar la estacionariedad en la media:

- La prueba de raíz unitaria Dickey–Fuller (DF).
- La prueba de raíz unitaria de Phillips–Perron (PP).

En el caso de que las pruebas estadísticas señalen que no hay estacionariedad en la media, se pueden realizar unas transformaciones llamadas diferenciaciones, las cuales se pueden realizar con un máximo de dos diferenciaciones para lograr una serie estacionaria en la media.

Si la serie de tiempo original resulta no estacionaria en la varianza luego de transformarla, los datos en estudio no pueden ser utilizados y por lo tanto se salen del programa. De manera similar, si a la serie de tiempo que siendo estacionaria en la varianza resulta no estacionaria en la media luego de diferenciarla dos veces, los datos en estudio no pueden ser utilizados y por lo tanto se salen del programa.

Para la realización de este procedimiento se utilizó la función llamada “adftest” ( $\text{adftest}(X)$ ) para la prueba de Dickey–Fuller y “pptest” ( $\text{pptest}(X)$ ) para la prueba Phillips–Perron, aplicada al vector de estudio.

#### 4.5. ESTUDIO DE AUTOCORRELACIÓN.



Finalizado el estudio de estacionariedad, la siguiente fase para asegurar resultados confiables es la determinación de la presencia o no de autocorrelación. Para este estudio se utilizan los siguientes contrastes de hipótesis:

- El contraste  $h$  de Durbin.
- El contraste de Durbin–Watson.

Para lograr que se rechace o no las hipótesis nulas de las pruebas estadísticas planteadas, los datos de la muestra deben seguir una regla de decisión bajo un nivel de significancia concreto.

El primer contraste a utilizar es el  $h$  de Durbin, ya que esta es una prueba con mayor confiabilidad, pero debido a que dicha prueba falla si el radicando es negativo, es decir, cuando  $[nVar(b_i) > 1]$  primero se verificará que no se cumpla dicha condición para poder aplicar la prueba, es decir, en los casos donde el radicando del estadístico de prueba del  $h$  de Durbin sea negativo, no se puede aplicar dicha prueba, y por consiguiente se aplicará el siguiente contraste.



La prueba de Durbin-Watson, arroja un estadístico que debe ser contrastado con un límite denominado "*du*" para determinar si los datos son o no autocorrelacionados.

El valor "*du*" depende del número de observaciones y de un valor *k* determinado, el cual representa el número de variables explicativas del proceso, dichos valores se encuentran tabulados, por lo que el programa debe invocar dicha tabla que se encuentra archivada en un documento de tipo ".xls" o "EXCEL 2003" que lleva por nombre "DurWat" y obtener el valor "*du*".

El programa contiene una función para el test de Durwin-Watson llamada "dwtest" ( `dwtest(X)` ).

Luego de aplicado el contraste conveniente, si la serie no presenta autocorrelación, se asume que los datos son independientes y por tanto pueden ser usados en el diseño de los gráficos de control y Análisis de punto de cambio. En caso contrario, los datos pasan a efectuarles un análisis de serie de tiempo por el método de Box y Jenkins.

#### 4.6. METODOLOGIA BOX & JENKINS.



El estudio consiste en graficar la función de autocorrelación y la función de autocorrelación parcial. De forma visual se analiza el comportamiento que presenta la serie y de acuerdo a unas reglas de decisión se concluye el modelo que mejor se aproxima a la serie.

#### 4.6.1. REGLA DE DECISIÓN PARA ELECCIÓN DE MODELOS DE SERIES DE TIEMPO.

- Si la Función de autocorrelación (FAC) presenta un decrecimiento rápido de tipo exponencial y / o sinusoidal la serie la aproxima a un modelo AR (p).
- Si la Función de autocorrelación Parcial (FACP) presenta un decrecimiento rápido de tipo exponencial y/o sinusoidal la serie la aproxima a un modelo MA (q).
- Si se cumple de forma conjunta las condiciones anteriores la serie se aproxima a un modelo ARMA (p, q).

Si la serie de tiempo se aproxima únicamente a un modelo MA (q), la serie no puede seguir en el estudio por tanto los datos salen del sistema. Esto se debe a que este modelo por sí solo no cumple con las propiedades necesarias para la siguiente etapa del programa.



Luego de determinado el modelo de serie de tiempo, se determinan los parámetros de la serie con las funciones apropiadas para tal fin.

Para poder obtener los parámetros cuando la serie se aproxima a un modelo ARMA (1,1), el usuario deberá utilizar el programa MINITAB, ya que la función que determina los parámetros en el lenguaje de programación MATLAB no determina todos los parámetros. El software diseñado le explicará detalladamente los pasos a seguir para la correcta obtención de dichos parámetros del modelo ARMA (1,1).

#### **4.7. GRÁFICOS DE CONTROL EWMA PARA DATOS NO AUTOCORRELACIONADOS.**

Para construir el gráfico de control EWMA, se necesita conocer de antemano los parámetros  $\lambda$  y  $k$ , con el fin de establecer los límites de control del gráfico, por tal motivo se debe comenzar realizando una estimación de ellos. Dichos parámetros se encuentran tabulados, sin embargo, para conseguirlos se necesita conocer el ARL que en este caso se utiliza un valor de 370 y un cambio en la media ( $D$ ), el cual debe ser seleccionado en el programa por el usuario. Una vez que se ha seleccionado el cambio en la media ( $D$ ) y con el ARL correspondiente de 370, el programa busca automáticamente en la tabla el valor de  $\lambda$  y  $k$ . La tabla para hallar los



valores de  $\lambda$  y  $k$  tabulados, se encuentran archivados en una tabla de tipo “.xls” o EXCEL 2003 y lleva por nombre “EWMA\_I”.

Una limitante importante es que la tabla solo proporciona valores de cambios de la media ( $D$ ) predeterminados, los cuales son: 0; 0.25; 0.5; 0.75; 1; 1.25; 1.5; 1.75; 2; 2.25; 2.5; 3; 3.25; 3.5; 3.75 y 4; por lo que el usuario deberá seleccionar alguno de estos valores y un ALR de 370 donde no tiene opción de cambiarlo.

Una vez obtenido el valor de  $\lambda$  y  $k$ , se procede de la siguiente manera para construir el gráfico de control:

- Se calculan los puntos que se van a graficar, es decir,  $Z_t$  (Promedios Móviles Ponderados Exponencialmente).
- Se calcula la media y la varianza de los  $Z_t$ .
- Se calculan los límites de control.
- Se construye el gráfico con los límites de control y los  $Z_t$ .

#### **4.8. GRÁFICOS DE CONTROL CUSUM PARA DATOS NO AUTOCORRELACIONADOS.**

Para construir los límites de control del gráfico CUSUM, únicamente se le pedirá al analista el número de desviaciones típicas que desea para detectar





cambios o corrimientos de esa magnitud. Dicho número lo escoge libremente el usuario, tomando en cuenta que debe ser mayor que cero y generalmente es menor que dos debido a que estos gráficos presentan un buen rendimiento frente a corrimientos pequeños de la media.

Luego de calculados los límites de control, y basados en ellos se calculan los vectores a graficar. Se elaboran dos, uno para las desviaciones positivas y otro para las desviaciones negativas. El vector de las desviaciones positivas será el vector determinado en la fórmula siempre que ese valor no sea menor que cero; si esto ocurre el valor en esa posición será cero. De forma inversa ocurre con el vector de las desviaciones negativas. Este será el vector determinado por las fórmulas siempre que sea menor que cero, si el valor en una posición es mayor que cero, el vector en esa posición valdrá cero.

En el caso de que algún valor del vector de las desviaciones positivas sea mayor que el límite de control superior, el programa arrojará un mensaje de que el gráfico se salió de control e inmediatamente calcula a partir de cual valor comenzó a ocurrir un cambio que produjo esta anomalía. De forma similar, si el vector de las desviaciones negativas es menor que el límite de control inferior, el programa arrojará un mensaje de que el gráfico se salió de control e inmediatamente calcula a partir de cual valor comenzó a ocurrir un



cambio que produjo esta anomalía. Finalmente el programa muestra el gráfico CUSUM correspondiente.

#### 4.9. GRÁFICOS DE CONTROL EWMA PARA DATOS AUTOCORRELACIONADOS.

Habiendo obtenido los parámetros del modelo por medio de la metodología Box y Jenkins se necesita obtener los parámetros  $\lambda$  y  $c$  para diseñar los límites de control y el vector  $Z$  (media móvil ponderado exponencialmente).

Estos valores están tabulados y se obtienen a partir de los valores de  $\Psi$ , de  $\phi$ , que son los parámetros del modelo, y un cambio en la media ( $\delta$ ) que deberá elegir el usuario arbitrariamente entre los valores siguientes: 0.5; 1; 1.5; 2; 2.5; 3; 4; 5, y buscando el menor valor de ARL se consigue el valor de  $\lambda$ .

Teniendo el valor de  $\lambda$  y los valores de  $\Psi$  y de  $\phi$ , el valor de  $c$  se obtiene a partir de otra tabla. Ver Tabla 4 en Apéndice I.

En el programa, la tabla que contiene los valores de  $\lambda$  se encuentra archivada en un documento de tipo “xls”, es decir, EXCEL 2003 de nombre



“EWMA\_A”. El usuario deberá invocar dicha tabla cuando el programa se le indique.

De la misma forma, la tabla que contiene los valores de  $c$  se encuentra archivada en un documento “.xls”, EXCEL 2003, de nombre “EWMA\_AC”.

Una limitante importante es que los valores de  $\Psi$  que podrán ser introducidos en la tabla son  $0.5$  o  $0.9$ , ya que son los únicos valores con que se construyeron dichas tablas. Se deberá elegir uno de los valores que más se parezca a los obtenidos en la metodología de Box y Jenkins. Así mismo, los valores de  $\phi$  que pueden ser introducidos en las tablas son  $0.4$  o  $0.8$ .

Una vez que se obtiene el valor de  $\lambda$  y  $c$ :

- Se calcula los puntos que se van a graficar, es decir,  $Z_t$  (Promedios Móviles Ponderados Exponencialmente).
- Se calcula la media de  $Z_t$ .
- Se calculan los límites de control.
- Con los límites de control y los  $Z_t$  se obtiene el gráfico.

#### 4.10. GRÁFICOS DE CONTROL CUSUM PARA DATOS AUTOCORRELACIONADOS.



Lo primero que se realiza es obtener los valores de  $r$  y  $c$  de las tablas correspondientes.

Al igual que en los gráficos EWMA para datos autocorrelacionados estos valores están tabulados y se obtienen a partir de los valores de  $\Psi$ , de  $\phi$ , que son los parámetros del modelo, y un cambio en la media ( $\delta$ ) que deberá elegir el usuario arbitrariamente entre los valores siguientes:  $0.5$ ;  $1$ ;  $1.5$ ;  $2$ ;  $2.5$ ;  $3$ ;  $4$ ; y  $5$ , y buscando el menor valor de ARL se consigue el valor de  $r$ . Ver tabla 5 en apéndice I.

Teniendo el valor de  $r$  y los valores de  $\Psi$  y de  $\phi$ , el valor de  $c$  se obtiene a partir de otra tabla. Ver Tabla 6 en Apéndice I.

En el programa, la tabla que contiene los valores de  $r$  se encuentra archivada en un documento de tipo “.xls”, es decir, “EXCEL 2003”, de nombre “CUSUM\_AU”. El usuario deberá invocar dicha tabla cuando el programa se le indique. De la misma forma, la tabla que contiene los valores de  $c$  se encuentra archivada en un documento “.xls”, EXCEL 2003, de nombre “CUSUM\_AC”.

Una limitante importante es que los valores de  $\Psi$  que podrán ser introducidos en la tabla son  $0.5$  o  $0.9$ , ya que son los únicos valores con que se construyeron dichas tablas. Se deberá elegir uno de los valores que más



se parezca a los obtenidos en la metodología de Box y Jenkins. Así mismo, los valores de  $\phi$  que pueden ser introducidos en las tablas son  $0.4$  o  $0.8$ . Habiendo obtenido los valores de  $r$  y  $c$  el procedimiento que se lleva a cabo para obtener el gráfico es similar al CUSUM para datos independientes. El único cambio es que ahora no se pedirá el número de desviaciones típicas, porque este valor viene representado por el valor de  $r$  ya determinado.

### ANÁLISIS DE PUNTO DE CAMBIO.

Hawkins y Zamba (2005) plantea una formulación para utilizarla en el control estadístico de procesos (SPC), donde el tamaño de la muestra no es fijo sino que crece indefinidamente, para ello usan límites de control ( $h_n$ ) en conjunto con la prueba estadística de Bartlett corregida ( $G_k, n$ ), la cual debe ser estimada a partir de todos los posibles puntos de separación  $k$  que se encuentren maximizando  $G_k, n$ , es decir, calculando  $G_{max}, n$ .

Para calcular el límite de control del análisis de punto de cambio, se le solicitará al usuario que indique el valor de alfa ( $\alpha$  es la probabilidad especificada de una señal errónea), los valores permitidos de  $\alpha$  son:  $0.001$ ;  $0.002$ ;  $0.005$ ;  $0.01$ ;  $0.02$ ;  $0.05$ . Una vez seleccionado el valor de alfa, y en el caso de que el número de observaciones sea mayor que  $14$ , se utiliza la aproximación dada para obtener los límites de control, de lo contrario, el programa invoca automáticamente una tabla de tipo “.xls” o



“EXCEL 2003” de nombre “TabHawkins”. Seguidamente el programa continúa con el cálculo del estimador  $G_{max},n$ , para verificar si dicho valor sobrepasa el límite de control. En el caso de que lo haga, el programa identifica que fue lo que exactamente cambió utilizando las pruebas estadísticas proporcionadas por MATLAB (Prueba  $t$  y Prueba  $F$  para dos muestras) para informar al analista si el cambio ocurrió en la media, en la varianza o en ambos.

El procedimiento en curso de SPC se plantea como sigue:

- Después de que la observación  $n$  se ha añadido al registro total del proceso, calcular  $G_{max},n$ .
- Si  $G_{max},n \leq h_n$ , donde  $h_n$  es un límite de control adecuado, entonces la conclusión es que no hay evidencia de un cambio en la media o varianza, y por lo tanto se debe dejar el proceso en ejecución ininterrumpida.
- Si  $G_{max},n > h_n$ , se concluye que hay evidencia de un cambio en la media, la varianza, o ambos.



## 5. APLICACIÓN DE LA HERRAMIENTA.

En el presente capítulo se ilustra el uso de la herramienta computarizada desarrollada, que a partir de casos prácticos reales de procesos industriales (en este caso data proveniente de la fabricación de papel) y casos teóricos, se evalúa su desempeño y funcionamiento al ponerlo a prueba con distintos tipos de datos, los cuales pueden ser o no estacionarios y presentar o no autocorrelación.

### **EJEMPLO N°1: Peso base ( $\text{g}/\text{m}^2$ ).**

El proceso de fabricación del papel higiénico usualmente se controla estadísticamente mediante el seguimiento univariante de las siguientes características de calidad: peso base, calibre y esfuerzo máximo a la tracción en la dirección de la máquina. El peso base ( $\text{g}/\text{m}^2$ ) representa la cantidad de material que se encuentra presente por cada  $\text{m}^2$  de papel producido. Para la medición del peso base se siguió la norma TAPPI T410 cm-02 y las lecturas se realizaron con una balanza analítica de 1 mg de precisión utilizando una probeta estándar.

En este ejemplo, se busca mostrar una serie de tiempo con distribución desconocida, la utilidad de la depuración de la data y su posterior utilización para la construcción de los gráficos de control adaptados a la misma.

**Tabla 5.1.** Datos de Peso Base en la fabricación de papel.

N°	Peso Base	N°	Peso Base	N°	Peso Base	N°	Peso Base
1	14,87	27	14,5	53	14,15	79	14,7
2	14,79	28	14,6	54	14	80	14,7
3	14,66	29	14,58	55	14,23	81	15
4	15,42	30	14,75	56	14,7	82	15,25
5	15	31	15	57	14,65	83	14,55
6	14,3	32	14,11	58	14,42	84	14,21
7	14,75	33	14,78	59	14,65	85	14,47
8	14,78	34	14,94	60	15,34	86	14,44
9	14,67	35	14,86	61	15,15	87	14,42
10	14,65	36	14,78	62	14,12	88	14,75
11	14,56	37	14,8	63	14,7	89	14,65
12	14,1	38	15,3	64	14,7	90	14,7
13	14,54	39	14,21	65	15,5	91	14,32
14	15	40	14,53	66	14,7	92	14,15





15	14,33	41	14,26	67	14,52	93	14,8
16	14,96	42	15,3	68	14,8	94	14,75
17	14,8	43	14,35	69	14,7	95	14,66
18	14,75	44	14	70	14,5	96	14,78
19	14,47	45	14,65	71	14,31	97	14,67
20	14,7	46	14,76	72	14,35	98	14,2
21	14,44	47	14,48	73	14,58	99	14,78
22	14,47	48	14,6	74	14,56	100	14,34
23	14,38	49	14,77	75	14,7	101	14,8
24	16,16	50	14,8	76	14,7	102	14,4
25	15,09	51	14,75	77	14,6	103	14,47
26	14,26	52	14,32	78	14,7	104	14,83

**Fuente:** Departamento control de calidad y ambiente (2012).

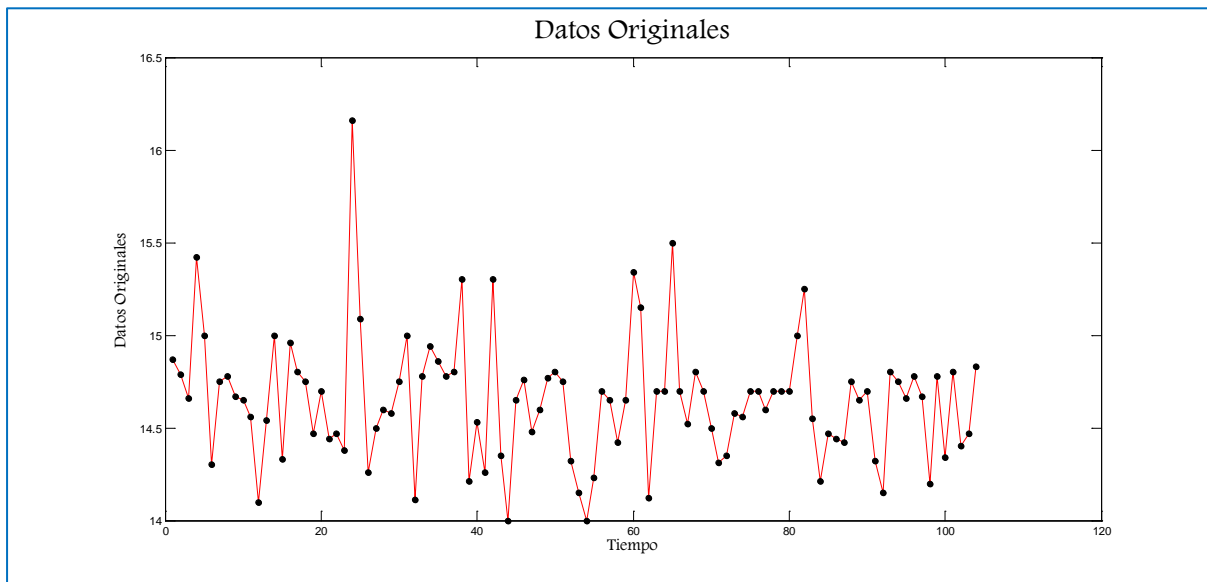
Al ejecutar el programa, aparece en pantalla la primera instrucción:

A continuación se debe leer el archivo que contiene los datos a estudiar para leerlos debe colocar el nombre del archivo EXCEL que contiene los datos a estudiar, hágalo cuando se le indique

Indique el nombre del archivo a leer: ejemplo\_1\_3



Una vez que el programa haya leído el nombre del archivo, muestra en pantalla el gráfico (en este caso la figura 5.1) para observar el comportamiento que presentan los datos originales:



**Figura 5.1.** Representación gráfica de los datos en su forma original.

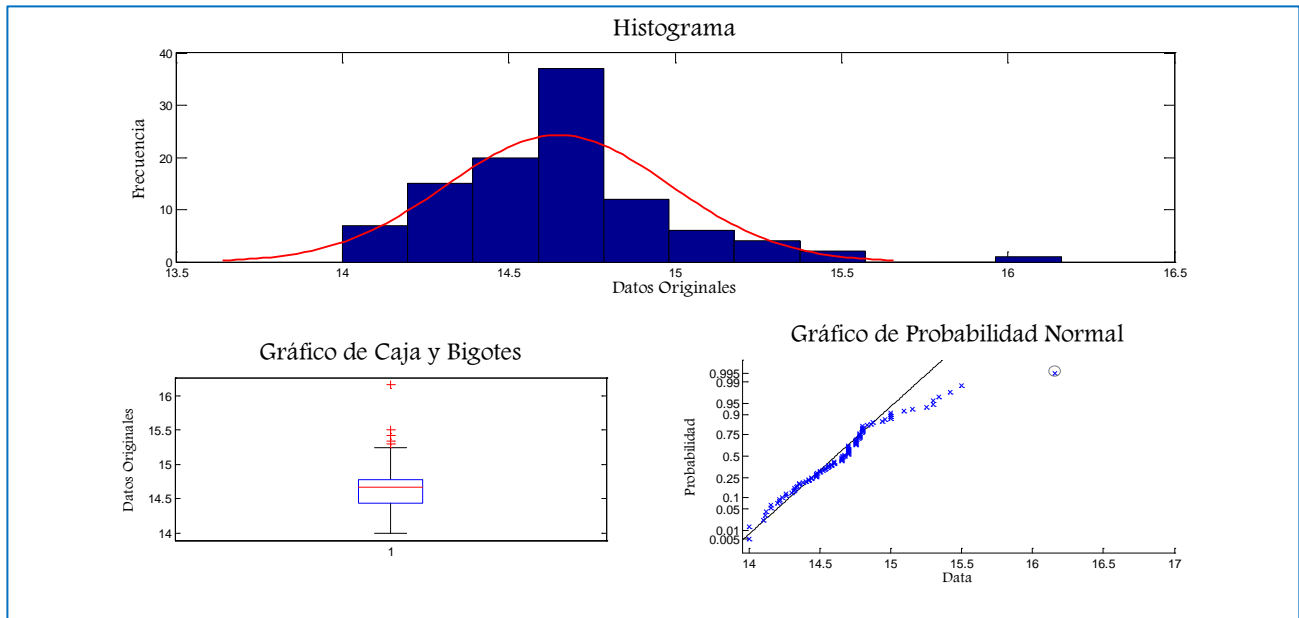


Figura 5.2. Representación gráfica de los datos en su forma original

Una vez observado el comportamiento de la data original, el programa muestra lo siguiente:

Los datos son aleatorios

Se realizó un contraste de rachas para comprobar independencia de los datos.

A continuación se realizará la eliminación de Outliers del vector original.

La detección se realiza mediante un proceso visual, la verificación de la presencia de los mismos y eliminación se realizó mediante la prueba de Rosner descrita en el capítulo II.

¿Autoriza la aplicación de dicha transformación?: Sí (1)  
No(0): 1



Indique el valor de alpha: 0.05

Este valor indica el nivel de confianza que deseo obtener para la aplicación de la prueba.

Indique el valor del número de outliers que espera encontrar: 5

Este valor es considerado por el usuario tomando en cuenta el origen de la muestra.

outliers detectados=

1

A continuación se muestra el comportamiento de los datos después de realizada la depuración de outliers:

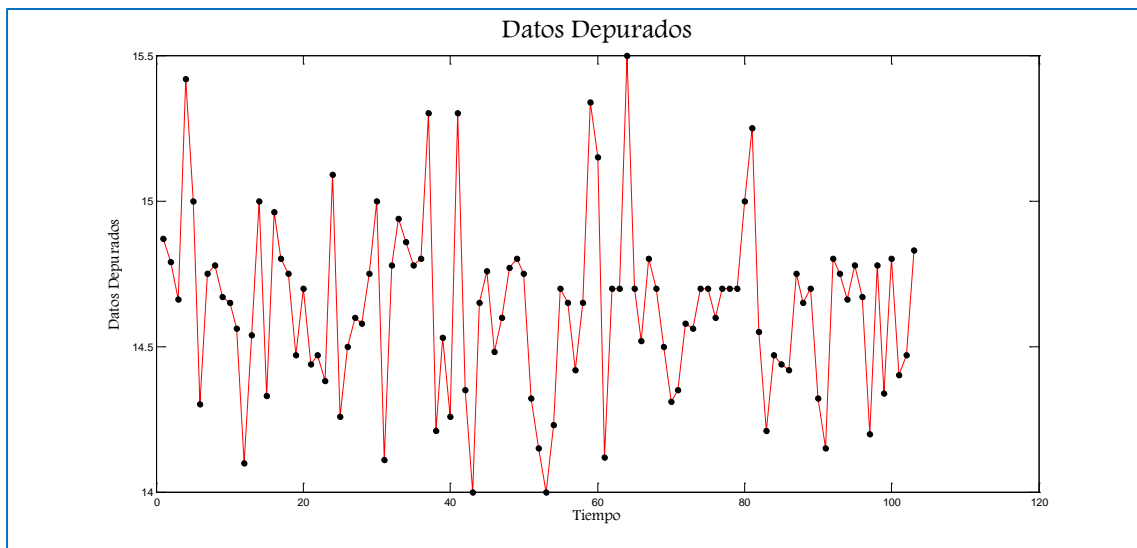


Figura 5.3. Representación gráfica de los datos depurados.

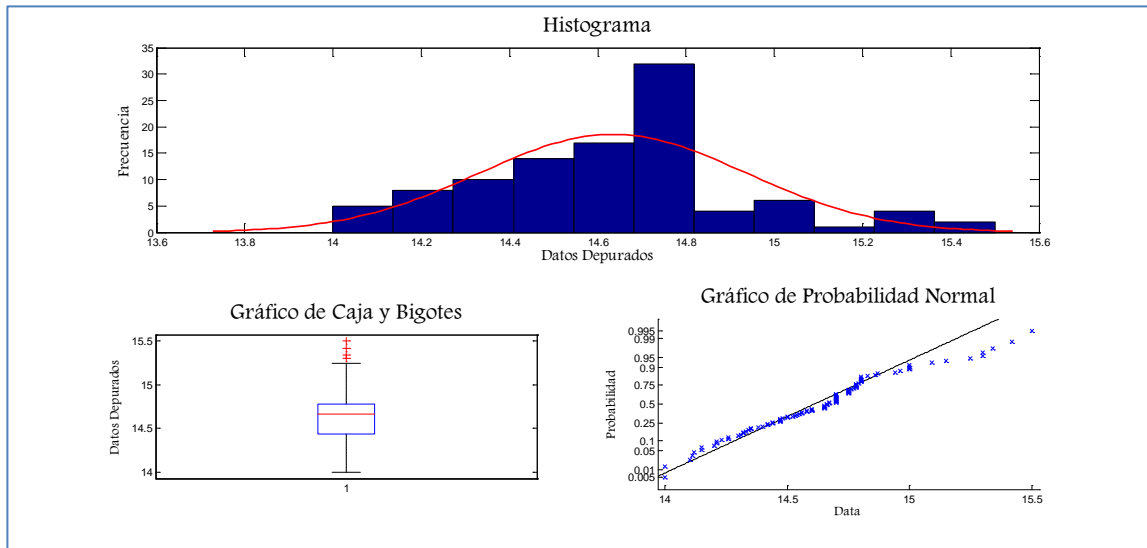


Figura 5.4. Representación gráfica de los datos depurados.

Al aplicar la prueba estadística de Lilliefors con un nivel de confianza del 95% (Alfa=5%), se rechaza la hipótesis que los datos proceden de una Distribución Normal, dado que p-valor del test realizado es inferior a 0.05

El valor de p es: 0.001

Esta función `lillietest(x)` prueba la hipótesis nula donde la muestra predeterminada en el vector `x` proviene de una distribución en la familia normal, frente a la alternativa de que no proceden de una distribución normal. La prueba devuelve el valor lógico `h = 1` si se rechaza la hipótesis nula al nivel de significación del 5%, y `h = 0` si no.



Se utiliza la transformación de Box-Cox para eliminar la no estacionariedad en la varianza, por lo que se debe autorizar al programa de que aplique dicha transformación. Seguidamente el programa muestra la siguiente instrucción:

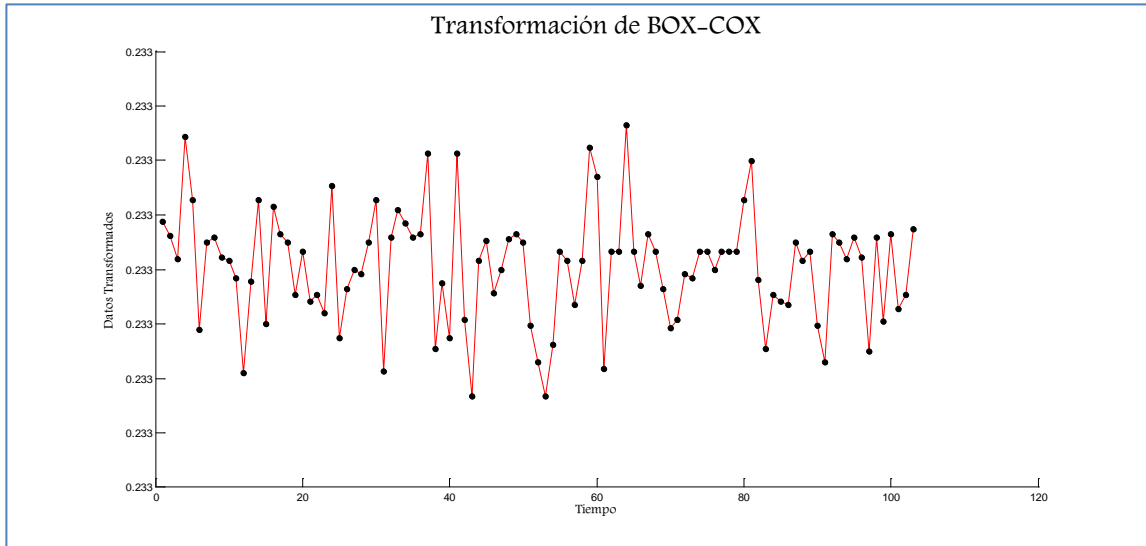
A continuación se realizará la transformación de BOX-COX que elimina la no estacionariedad en la varianza.

¿Autoriza la aplicación de dicha transformación?: Si(1)  
No(0) 1

¿Desea graficar la serie estacionaria en cuanto a la  
Varianza?: Si(1) No(0) 1

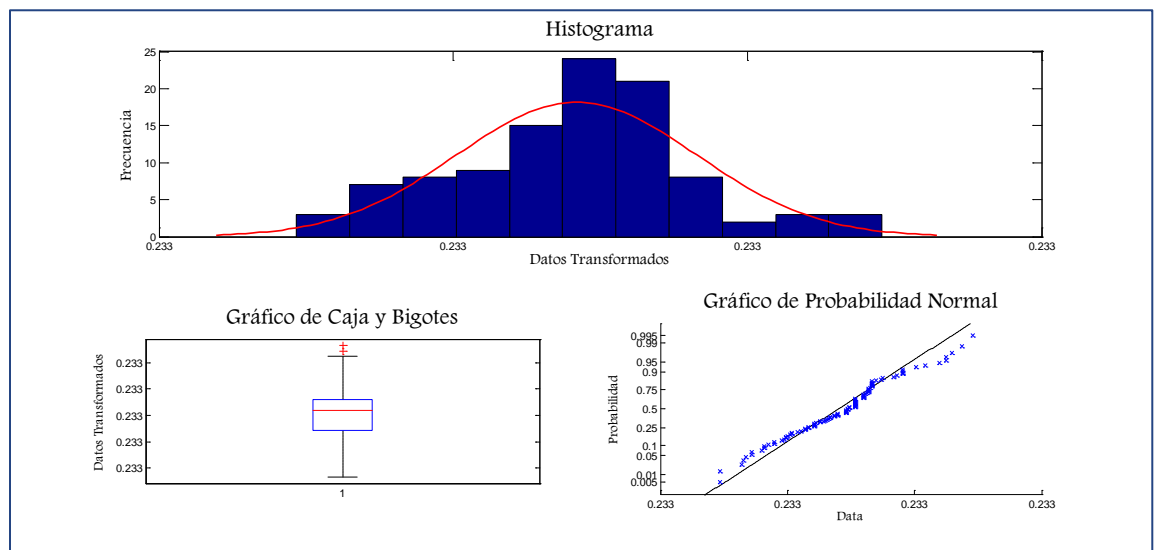
Después de realizada esta transformación se le permite al usuario la visualización de los valores transformados para observar los cambios en el proceso.

Al aceptarla muestra en pantalla la figura 5.5 y en el caso contrario continúa con el siguiente paso que corresponde a evaluar la estacionariedad en la media:



**Figura 5.5.** Representación gráfica de los datos al aplicarles la transformación de Box-Cox.

En la figura 5.5 se observa que disminuye la dispersión con respecto a la que presentaban los datos depurados.



**Figura 5.6.** Representación gráfica de los datos transformados.



Como se verá a continuación, el programa aplica las pruebas estadísticas automáticamente y el resultado de la prueba indica que los datos no son estacionarios en la media, estas transformaciones provienen de la aplicación de las pruebas de raíz unitaria DF y PP por lo que es necesario autorizar al programa para que realice las transformaciones llamadas “diferenciaciones” para lograr la estacionariedad en la media. Se comienza con la primera diferenciación como sigue:

Los datos no son estacionarios en su forma original.

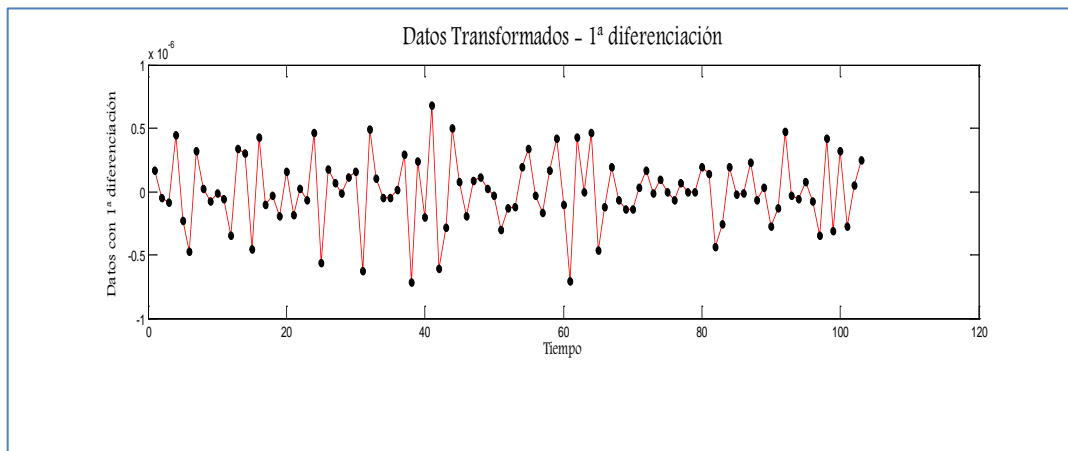
¿Autoriza la realización de una 1º diferenciación?:

Si(1) No(0) 1

¿Desea observar los datos diferenciados 1º vez?: Si(1)

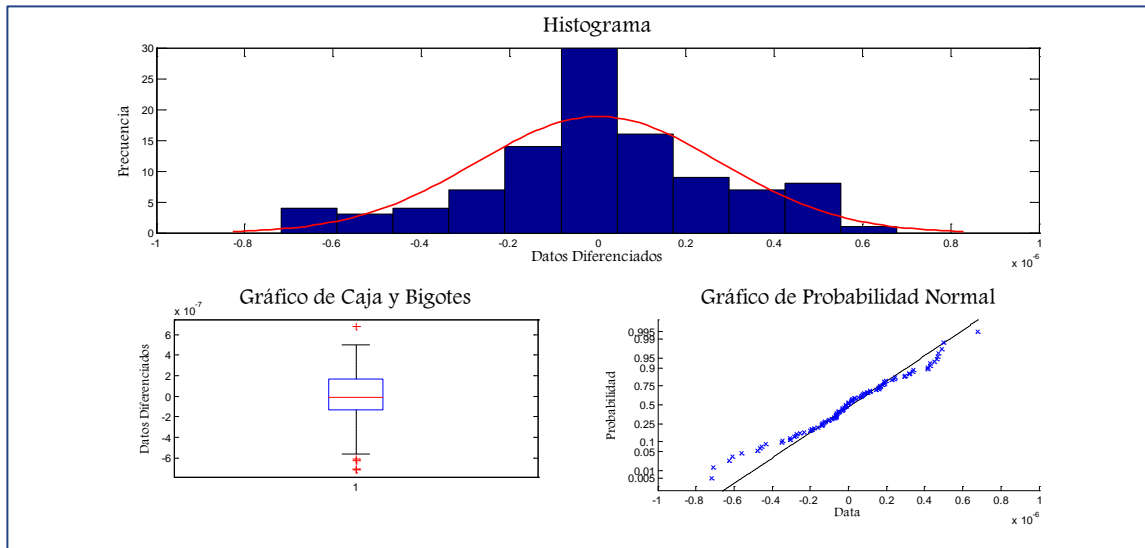
No(0) 1

Los datos son estacionarios en 1º diferenciación



**Figura 5.7.** Representación gráfica de los datos al aplicarles la primera diferenciación.





**Figura 5.8.** Representación gráfica de los datos al aplicarles la primera diferenciación.

Posteriormente se realizó el estudio de autocorrelación mediante las pruebas h-Durbin o Durbin-Watson.

A continuación se aplicará la prueba H-DURBIN para determinar si los datos son o no autocorrelacionados.

¿Autoriza la realización de esta prueba?: Si(1) No(0) 1  
Los datos sometidos a evaluación presentan autocorrelación

Este mensaje es enviado para que el usuario decida si desea trabajar con datos autocorrelacionados.

Los datos presentan autocorrelación.

¿Desea ignorar el fenómeno de autocorrelación?: Si(1)  
No(0) 1

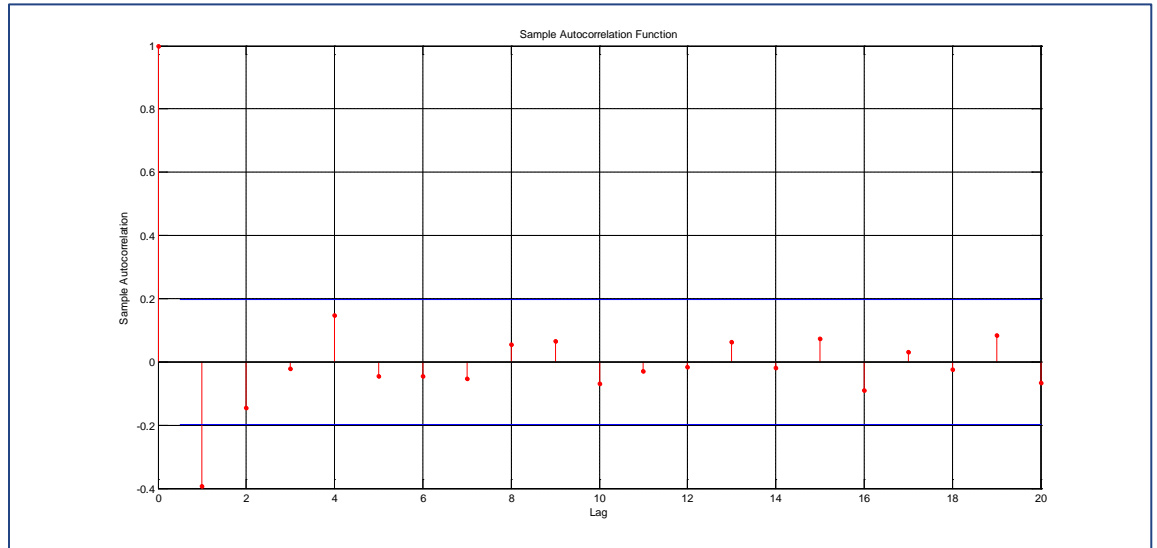


En este momento el usuario decide si ignora el fenómeno de autocorrelación y continuar realizando los Gráficos de Control CUSUM y EWMA para datos Autocorrelacionados:

A continuación se mostrará el gráfico de los coeficientes de autocorrelación

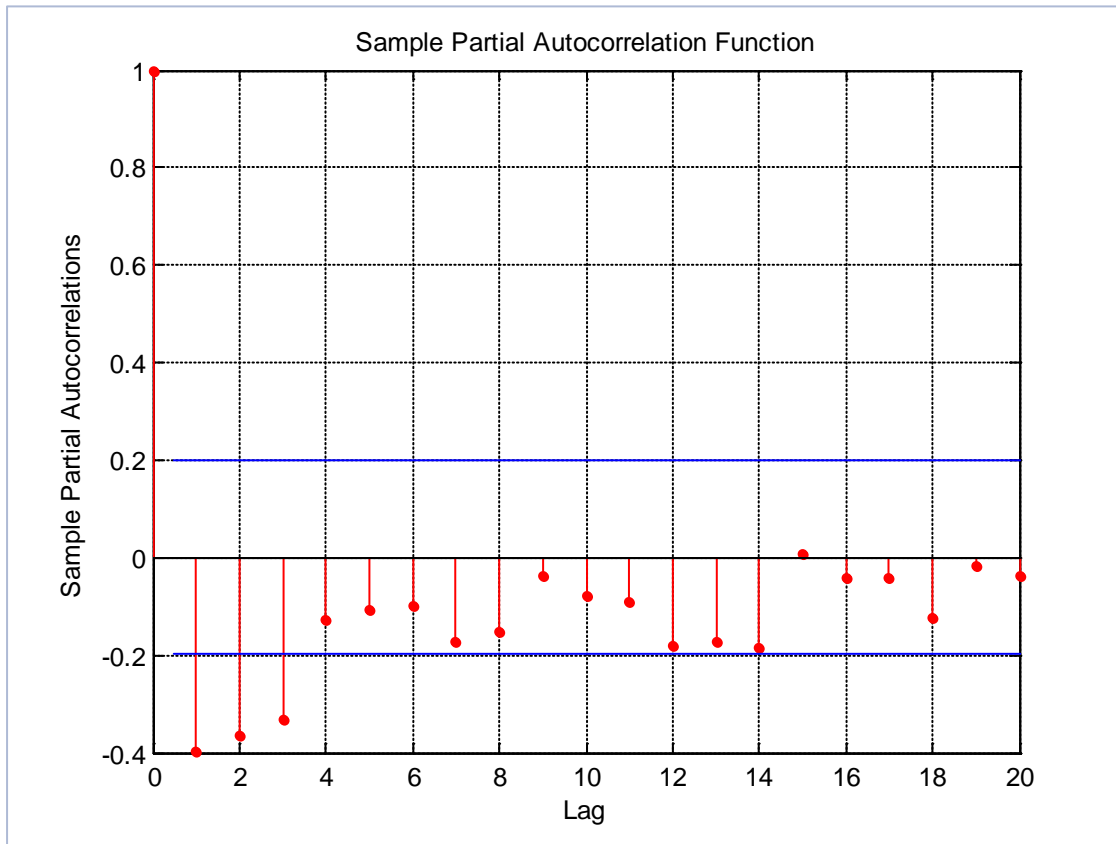
Autoriza mostrar dicha gráfica? Si(1) No(0):1

La visualización de este gráfico permite observar el decrecimiento o no de coeficientes de autocorrelación y así determinar el modelo que sigue la serie de tiempo.



**Figura 5.9.** Representación gráfica de los coeficientes de autocorrelación

¿Los coeficientes descienden rapidamente de forma exponencial y/o sinusoidal? Si(1) No(0): 1



**Figura 5.10.** Representación gráfica de los coeficientes de autocorrelación parcial.

¿Los coeficientes descienden rápidamente de forma exponencial y/o sinusoidal?: 0

La serie sigue un proceso AR

Dado el valor de  $\phi_1$  de las formulas se debe elegir

El valor tabulado más cercano, el valor real es:

$\phi_1 =$



0.3955

PUEDE ELEGIR  $\Phi = 0.4$  ó  $0.8$

DIGA EL VALOR DE  $\Phi$ : 0.4

$\Phi =$

0.4000

Dado el valor de  $Sh_{i1}$  de las formulas se debe elegir

El valor tabulado más cercano, el valor real es:

$Sh_{i1} =$

8.3839e-042

PUEDE ELEGIR  $Sh_i = 0.5$  ó  $0.9$

DIGA EL VALOR DE  $Sh_i$ : 0.5

$Sh_i =$

0.5000

A continuación se observa la media de los datos transformados

MEDIAA =

14.6344



Debe elegir entre los siguientes valores que están  
Tabulados el valor del cambio en la media que usted quiere  
Detectar más rápidamente

0,5    1        1,5        2        2,5        3        4        5

INDIQUE EL VALOR SELECCIONADO: 2

cambt =

2

A continuación deberá leer el archivo Excel 2003 que  
contiene los valores De lamda en función de Phi, Shi y el  
cambio en la media

Para leer el archivo deberá colocar el nombre "EWMA\_A"  
cuando se le indique

Cuide colocar el nombre en mayúsculas y colocar el guion  
bajo

Indique el nombre de archivo a leer: EWMA\_A

filename =

EWMA\_A

lamdat =

0.3500

Debe elegir entre los siguientes valores que están



Tabulados el valor lamda que sea mas cercano al lamda obtenido

Para efectos de buscar en la tabla de los C

0.1      0.2      0.3

INDIQUE EL VALOR SELECCIONADO: 0.3

LAMDA =

0.3000

DADO EL VALOR DE  $\Phi_{il}$  DE LAS FÓRMULAS SE DEBE ELEGIR EL VALOR TABULADO MÁS CERCANO, EL VALOR REAL ES:

$\Phi_{il}$  =

0.3955

PUEDE ELEGIR  $\Phi_i$  = 0.1   0.2   0.3   0.4   0.5   0.6   0.7   0.8  
0.9

DIGA EL VALOR DE  $\Phi_i$ : 0.4

$\Phi_i$  =

0.4000

DADO EL VALOR DE  $\Phi_{i1}$  DE LAS FÓRMULAS SE DEBE ELEGIR EL VALOR TABULADO MÁS CERCANO, EL VALOR REAL ES:

$\Phi_{i1}$  =



8.3839e-042

PUEDE ELEGIR Shi = 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8  
0.9

DIGA EL VALOR DE Shi: 0.1

Shi =

0.1000

A CONTINUACIÓN DEBERÁ LEER EL ARCHIVO EXCEL QUE CONTIENE  
LOS VALORES

DE C EN FUNCIÓN DE Phi, Shi Y DE LAMDA

PARA LEER EL ARCHIVO DEBERÁ COLOCAR EL NOMBRE "EWMA\_AC"  
CUANDO SE LE INDIQUE

CUIDE COLOCAR EL NOMBRE EN MAYÚSCULAS Y COLOCAR EL GUION  
BAJO

Indique el nombre de archivo a leer: EWMA\_AC

filename =

EWMA\_AC

DIGA true PARA MOSTRAR EL GRÁFICO DE CONTROL EWMA

AUTOCORRELACIONADO: true

grafewa =





1

A continuación el programa presenta el gráfico de control EWMA para datos autocorrelacionados:

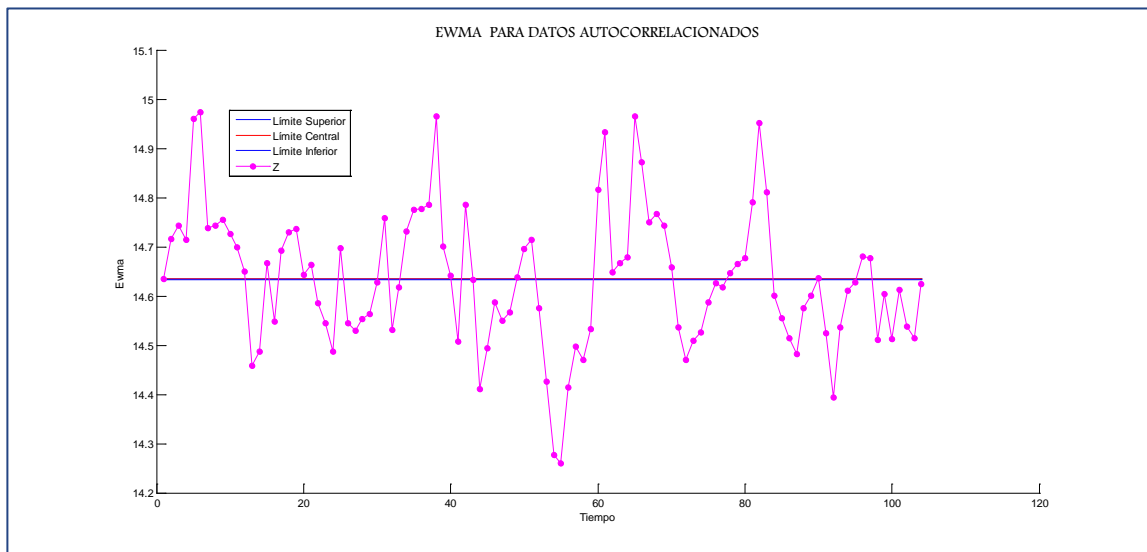


Figura 5.11. Representación gráfica EWMA para datos autocorrelacionados.

Se observa que el proceso no esta bajo control.

Dado el valor de  $\Phi$  de las fórmulas se debe elegir el valor tabulado más cercano, el valor real es:

$\Phi =$

0.3955

Puede elegir  $\Phi = 0.4$  ó  $0.8$

Diga el valor de  $\Phi$ : 0.4



Phi =

0.4000

Dado el valor de  $Sh_{i1}$  de las fórmulas se debe elegir  
El valor tabulado más cercano, el valor real es:

$Sh_{i1}$  =

8.3839e-042

Puede elegir  $Sh_i = 0.5$  ó  $0.9$

Diga el valor de  $Sh_i$ : 0.5

$Sh_i$  =

0.5000

A continuación vera la media de los datos transformados

MEDIAA =

14.6344

Debe elegir entre los siguientes valores que estan  
tabulados el valor del cambio en la media que usted  
quiere

detectar más rápidamente



0,5    1        1,5        2        2,5        3        4        5

Indique el valor seleccionado: 1.5

cambt =

1.5000

A continuación deberá leer el archivo EXCEL 2003 que contiene los valores

de LAMDA en función de Phi, Shi y el cambio en la media para leer el archivo deberá colocar el nombre

"CUSUM\_AU" cuando se le indique

cuide colocar el nombre en mayúsculas y colocar el guión bajo

Indique el nombre de archivo a leer: CUSUM\_AU

filename =

CUSUM\_AU

erre =

0.7000

Debe elegir entre los valores que estan tabulados el valor r que sea más cercano al erre obtenido

para efectos de buscar en la tabla de los C



0.25      0.5      1

Indique el valor seleccionado: 1

errel =

1

Dado el valor de Phi1 de las fórmulas se debe elegir el valor tabulado más cercano, el valor real es:

Phi1 =

0.3955

Puede elegir Phi = 0.1    0.2    0.3    0.4    0.5    0.6  
0.7    0.8    0.9

Diga el valor de Phi: 0.4

Phi =

0.4000

Dado el valor de Shil de las fórmulas se debe elegir El valor tabulado más cercano, el valor real es:

Shil =

8.3839e-042



Puede elegir Shi = 0.1 0.2 0.3 0.4 0.5 0.6  
0.7 0.8 0.9

Diga el valor de Shi: 0.1

Shi =

0.1000

A continuación deberá leer el archivo EXCEL 2003 que contiene los valores

de LAMDA en función de Phi, Shi y el cambio en la media

para leer el archivo deberá colocar el nombre "CUSUM\_AC" cuando se le indique

cuide colocar el nombre en mayúsculas y colocar el guión bajo

Indique el nombre de archivo a leer: CUSUM\_AC

filename =

CUSUM\_AC

El siguiente valor indica la posición en la que los datos comenzaron a estar

Fuera\_de\_Controla =



El siguiente valor indica la posición en la que  
ocurrio un

cambio =

2

Diga true para mostrar el gráfico CUSUM para datos  
autocorrelacionados: true

grafcusa =

1

A continuación el programa presenta el gráfico de control CUSUM para  
datos autocorrelacionados:

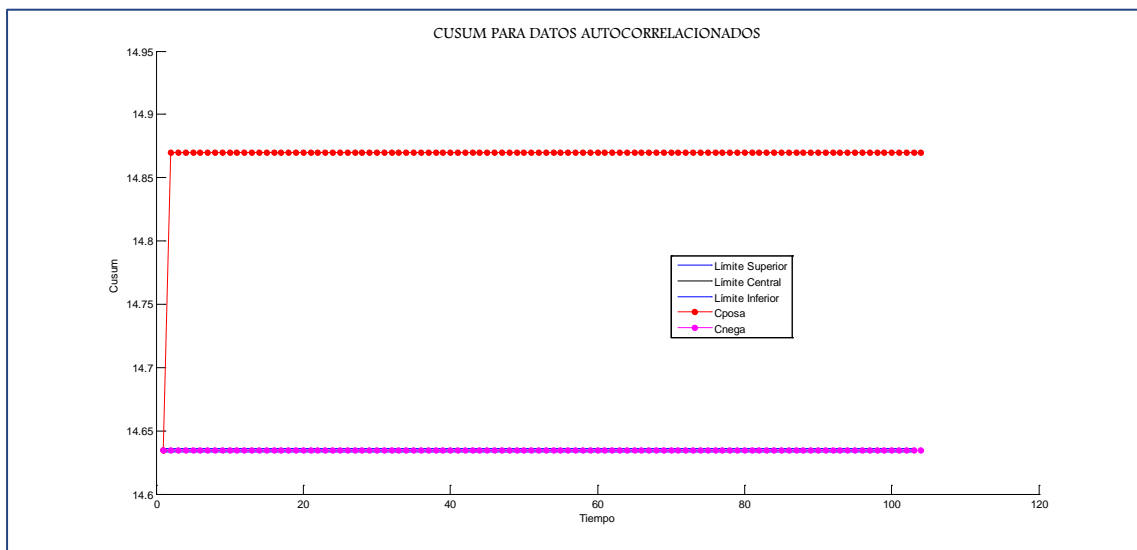


Figura 5.12. Representación gráfica de CUSUM para datos autocorrelacionados.



Se observa que el cambio ocurre en la posición 2 y a partir de ese valor el proceso esta fuera de control.

**EJEMPLO N°2: Datos de tamaño de partícula ( $\mu\text{m}$ ).**

En este ejemplo, se busca ilustrar la utilidad del análisis exploratorio de datos, en cuanto a la variación de comportamiento antes y después de la depuración y su posterior construcción de los gráficos de control.

Una planta elabora un producto en forma de polvo. Un parámetro importante es el tamaño promedio de partícula del producto, que se mide con un analizador automático. Se selecciona de forma aleatoria una muestra del producto de cada lote de producción, se mezcla bien y se divide en dos mitades que se enumeran como N°1 y N°2 antes de someterse al analizador de partículas. Los resultados obtenidos para 31 lotes consecutivos se muestran en la siguiente tabla:



**Tabla 5.2.** Datos de tamaño de partículas

N°	Tamaño 1 ( $\mu\text{m}$ )	Tamaño 2 ( $\mu\text{m}$ )
1	180	191
2	145	148
3	174	178
4	153	162
5	143	156
6	162	172
7	181	185
8	177	191
9	152	166
10	165	178
11	162	162
12	169	173
13	147	149
14	168	168
15	152	166
16	189	200





17	155	169
18	145	151
19	164	172
20	180	182
21	171	180
22	156	161
23	161	163
24	169	169
25	155	164
26	167	171
27	160	161
28	144	156
29	174	177
30	184	194
31	536	471

**Fuente:** Gutiérrez y De La Vara (2004).

Al ejecutar el programa, aparece en pantalla la primera instrucción:

A continuación debe leer el archivo que contiene los datos a estudiar, para leerlos debe colocar el nombre del archivo EXCEL que contiene los datos a estudiar, hágalo cuando se le indique

Indique el nombre del archivo a leer: tamaño.xlsx

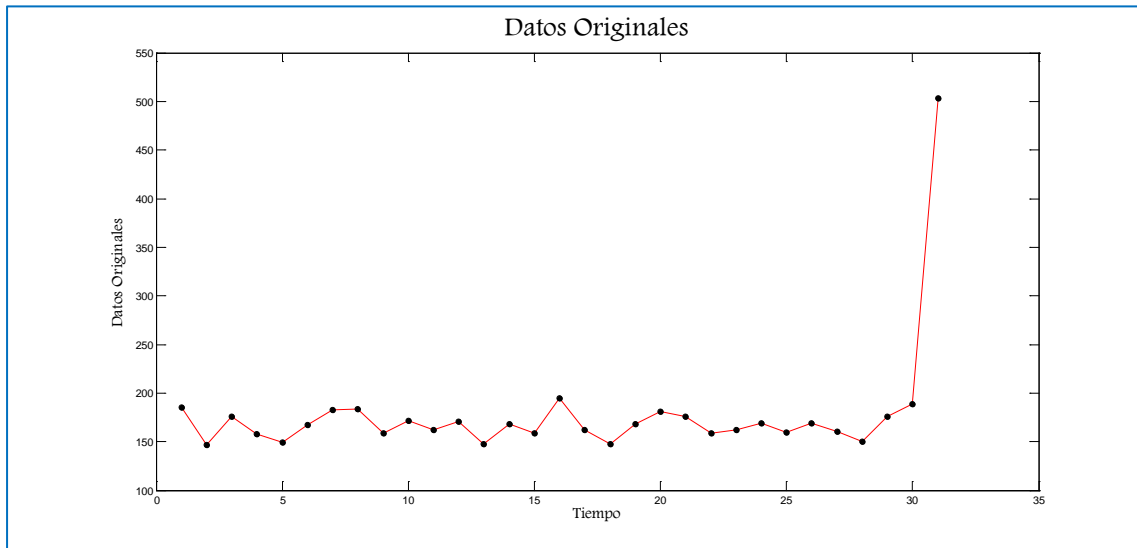


Figura 5.13. Representación gráfica de los datos en su forma original.

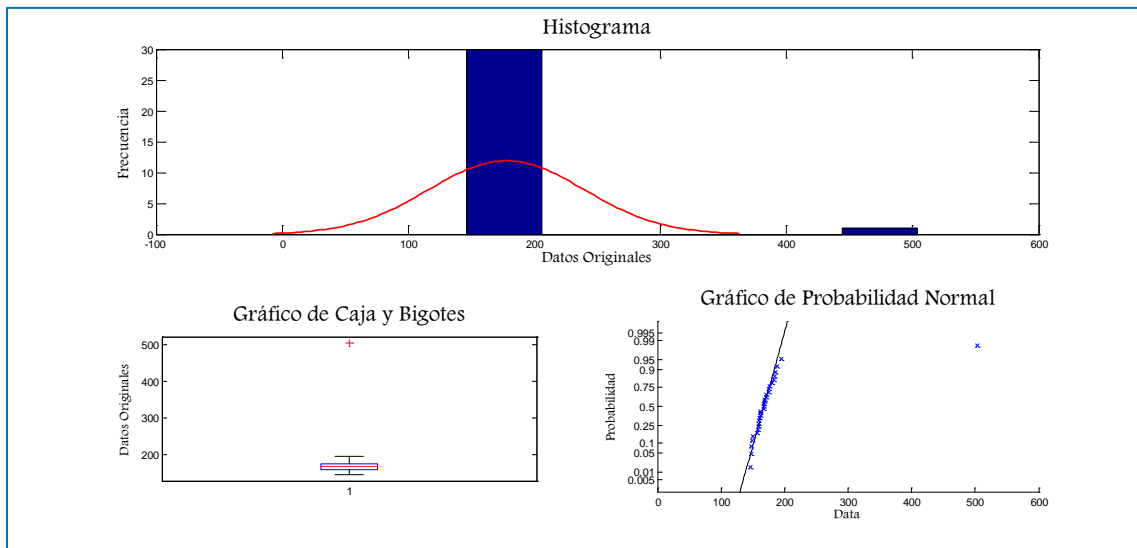


Figura 5.14. Representación gráfica de los datos en su forma original.

Los datos son aleatorios



A continuación se realiza la depuración de la data suministrada:

A continuación se realizará la eliminación de Outliers del vector original.

¿Autoriza la aplicación de dicha transformación?: Si(1)

No(0) 1

Indique el valor de alpha: 0.05

Indique el valor del número de outliers que espera encontrar: 6

outliers detectados=

1

Al aplicar la prueba estadística de Lilliefors con un nivel de confianza del 95% (Alfa=5%), no hay suficiente evidencia estadística para rechazar la hipótesis que los datos proceden de una Distribución Normal; dado que el p-valor del test realizado es mayor a 0.05

El valor de p es: 0.34243

Advertencia: Se recomienda el uso del gráfico EWMA por ser robusto en la presencia de distribuciones no Normales, a diferencia del CUSUM y Análisis de punto de cambio.

A continuación se realizará la transformación de BOX-COX que elimina la no estacionariedad en la varianza.



¿Autoriza la aplicación de dicha transformación?: Si(1)  
No(0)1

Representación gráfica de los datos después de la eliminación de outliers:

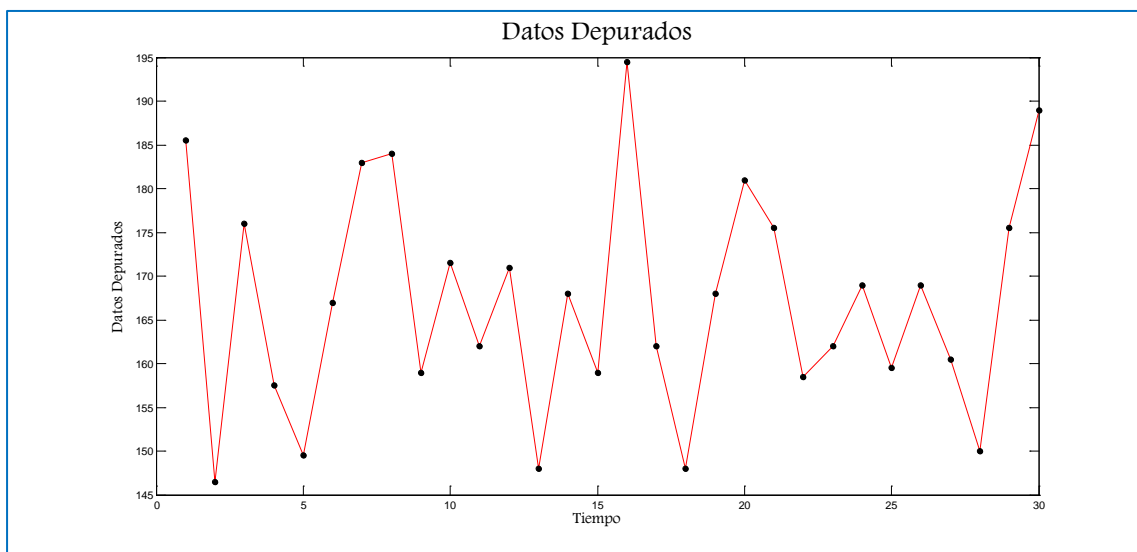


Figura 5.15. Representación gráfica de los datos depurados.

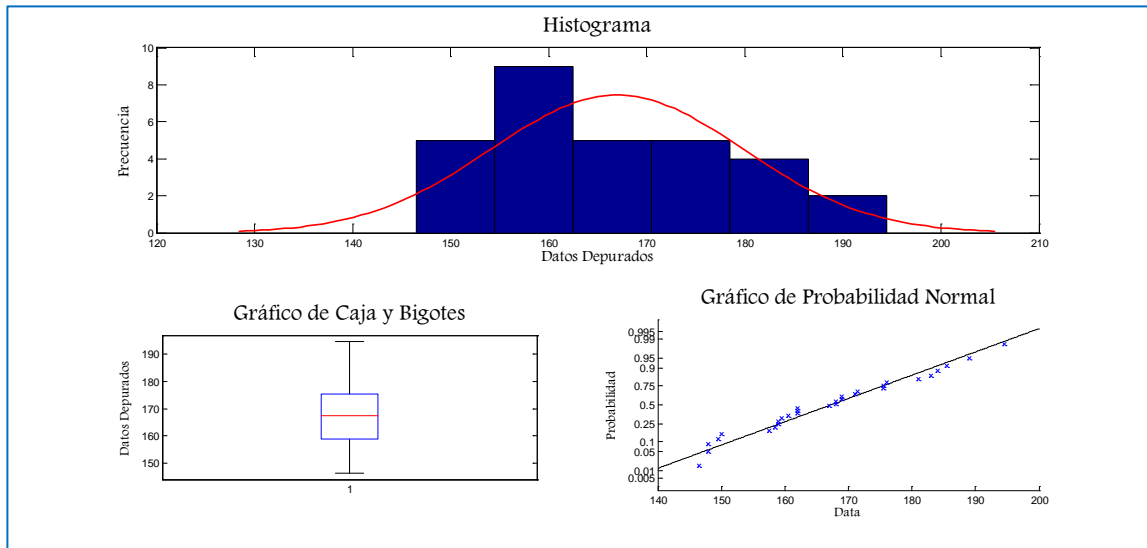


Figura 5.16. Representación gráfica de los datos depurados.

Al aplicar la prueba estadística de Lilliefors con un nivel de confianza del 95% ( $\alpha=5\%$ ), no hay suficiente evidencia estadística para rechazar la hipótesis que los datos proceden de una Distribución Normal; dado que el p-valor del test realizado es mayor a 0.05

El valor de p es: 0.34243

Advertencia: Se recomienda el uso del gráfico EWMA por ser robusto en la presencia de distribuciones no Normales, a diferencia del CUSUM y Análisis de punto de cambio.



A continuación se realizará la transformación de BOX-COX que elimina

la no estacionariedad en la varianza.

¿Autoriza la aplicación de dicha transformación?: Si(1)

No(0) 1

¿Desea graficar la serie estacionaria en cuanto a la

Varianza?: Si(1) No(0) 1

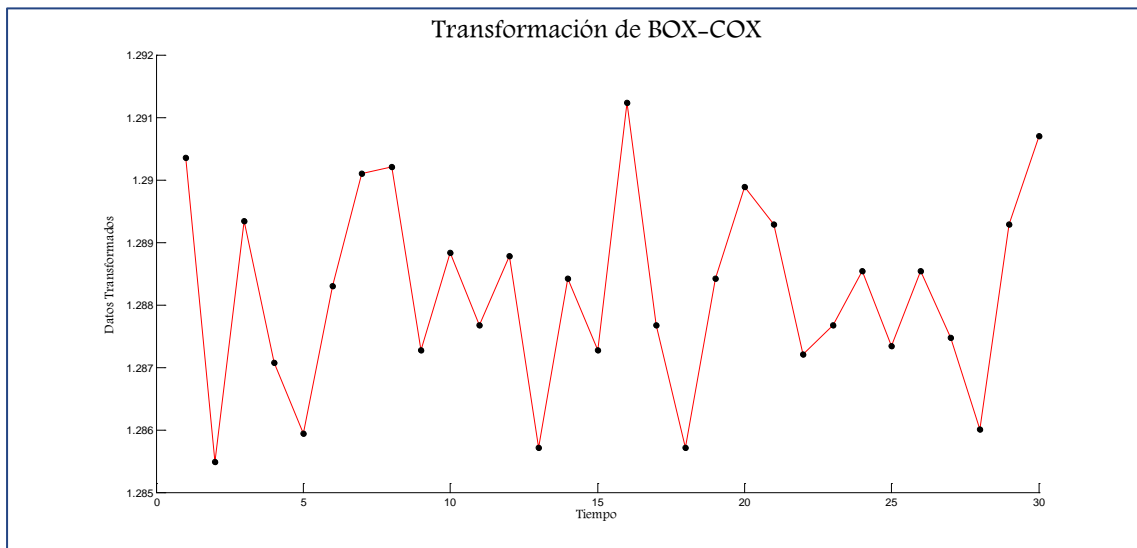


Figura 5.17. Representación gráfica de los datos transformados.

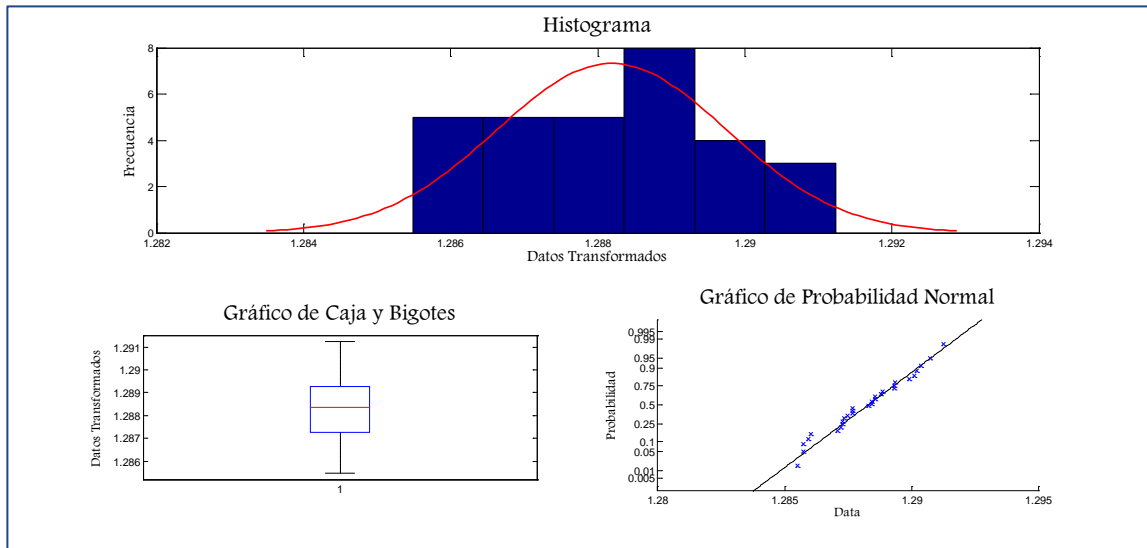


Figura 5.18. Representación gráfica de los datos transformados.

Los datos no son estacionarios en su forma original.

¿Autoriza la realización de una 1ª diferenciación?: Si(1)

No(0) 1

¿Desea observar los datos diferenciados 1ª vez?: Si(1)

No(0) 1

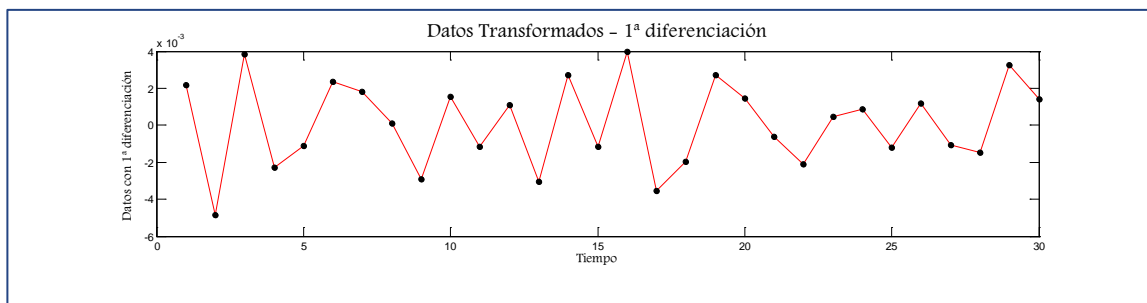
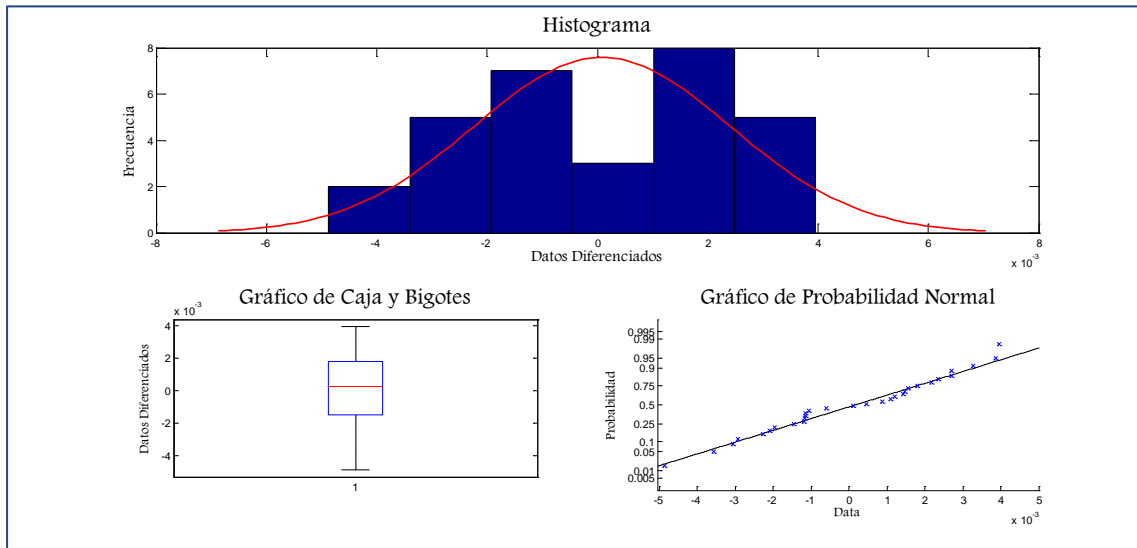


Figura 5.19. Representación gráfica de los datos diferenciados.



**Figura 5.20.** Representación gráfica de los datos diferenciados.

Los datos son estacionarios en 1ª diferenciación

A continuación se aplicará la prueba h-DURBIN para determinar si los datos son o no autocorrelacionados.

¿Autoriza la realización de esta prueba?: Si(1) No(0) 1

Los datos sometidos a evaluación presentan autocorrelación

¿Desea ignorar el fenómeno de autocorrelacion?: Si(1) No(0)

0

ignauto =0

**El programa se sale del sistema.**

**EJEMPLO N°3: Peso base (g/m<sup>2</sup>)**





El proceso de fabricación del papel higiénico usualmente se controla estadísticamente mediante el seguimiento univariante de las siguientes características de calidad: peso base, calibre y esfuerzo máximo a la tracción en la dirección de la máquina. El peso base ( $\text{g/m}^2$ ) representa la cantidad de material que se encuentra presente por cada  $\text{m}^2$  de papel producido. Para la medición del peso base se siguió la norma TAPPI T410 cm-02 y las lecturas se realizaron con una balanza analítica de 1 mg de precisión utilizando una probeta estándar.

En este ejemplo, se busca mostrar una serie de tiempo con distribución desconocida, la utilidad de la depuración de la data y su posterior utilización para la construcción de los gráficos de control adaptados a la misma.

**Tabla 5.3** Datos de Peso Base en la fabricación de papel.

N°	Peso Base	N°	Peso Base	N°	Peso Base	N°	Peso Base
1	14,57	27	14,7	53	14,7	79	15
2	14,75	28	14,6	54	14,3	80	14,65
3	14,42	29	14,7	55	14,55	81	14,6
4	14,75	30	14	56	14,71	82	14,7
5	14,57	31	14,7	57	15	83	15
6	15	32	14,6	58	14,78	84	14,68
7	14,59	33	14	59	14,51	85	15,36
8	15,15	34	14,6	60	14,62	86	15
9	15,15	35	14,78	61	14,49	87	14,9
10	14,62	36	14,59	62	14	88	14,15
11	14,78	37	14,63	63	14,7	89	14
12	14,85	38	14,6	64	15	90	14,6
13	14,47	39	14,9	65	14,43	91	14,58
14	14,4	40	14,75	66	14,7	92	14,56
15	14,8	41	14,8	67	14,38	93	14,8
16	15	42	15	68	14,7	94	14,65
17	14,78	43	14,7	69	14,8	95	14,75
18	15,1	44	14,8	70	14,6	96	15
19	14,75	45	15	71	14,8	97	14,6
20	14,98	46	14,6	72	14,72	98	14,75



21	14,88	47	14,4	73	14,57	99	14,75
22	15	48	14,78	74	15,4	100	14,3
23	15,5	49	14,4	75	14,56	101	14,65
24	14,7	50	14,8	76	14,2		
25	14,39	51	15	77	14,65		
26	14,9	52	14,7	78	14,8		

**Fuente:** Departamento control de calidad y ambiente (2012).

Al ejecutar el programa, aparece en pantalla la primera instrucción:

A continuación debe leer el archivo que contiene los datos a estudiar

para leerlos debe colocar el nombre del archivo EXCEL que contiene

los datos a estudiar, hágalo cuando se le indique

Indique el nombre del archivo a leer: ejemplo\_1\_10.xlsx

Los datos son aleatorios

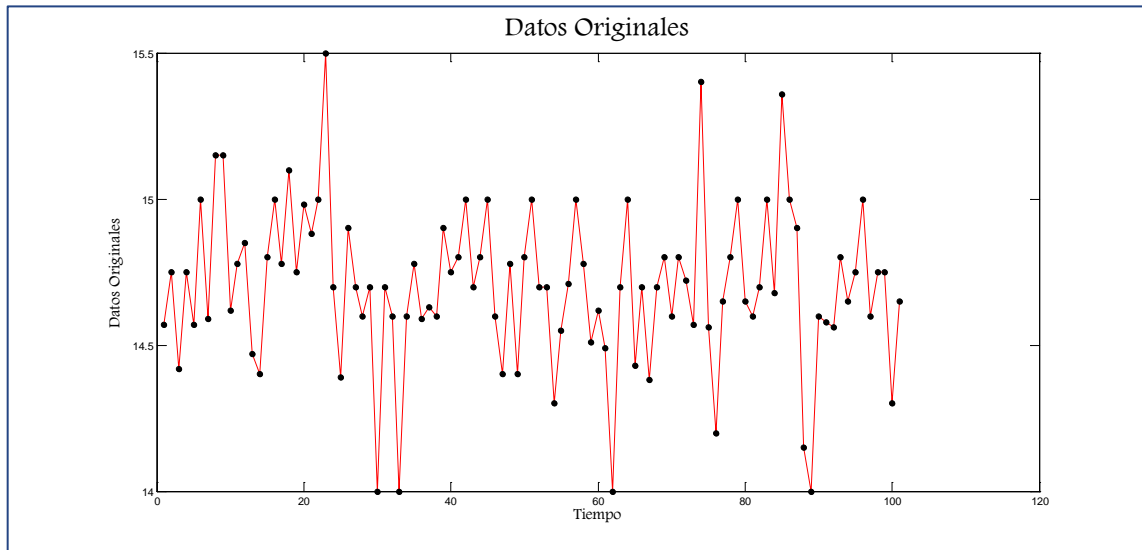


Figura 5.21. Representación gráfica de los datos en su forma original.

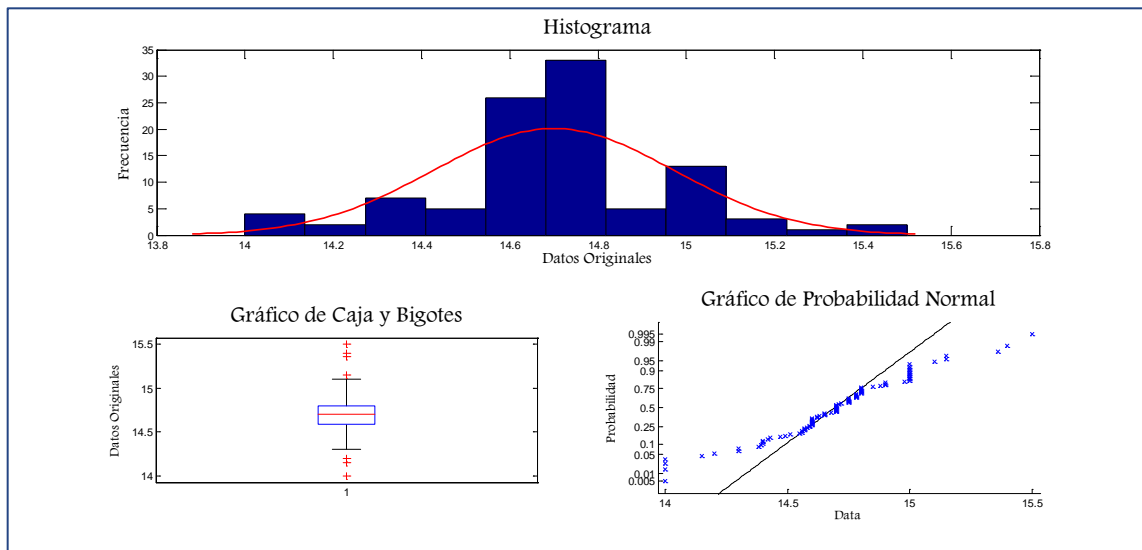


Figura 5.22. Representación gráfica de los datos en su forma original.

A continuación se realizará la eliminación de Outliers del vector original.



¿Autoriza la aplicación de dicha transformación?: Si(1)  
No(0) 1

Indique el valor de alpha: 0.05

Indique el valor del número de outliers que espera encontrar: 5

outliers detectados=

0

Al aplicar la prueba estadística de Lilliefors con un nivel de

confianza del 95% (Alfa=5%), se rechaza la hipótesis que los datos proceden

de una Distribución Normal, dado que p-valor del test realizado es inferior a 0.05

El valor de p es: 0.0012758

A continuación se realizará la transformación de BOX-COX que elimina

la no estacionariedad en la varianza.

¿Autoriza la aplicación de dicha transformación?: Si(1)  
No(0) 1

¿Desea graficar la serie estacionaria en cuanto a la Varianza?: Si(1) No(0) 1

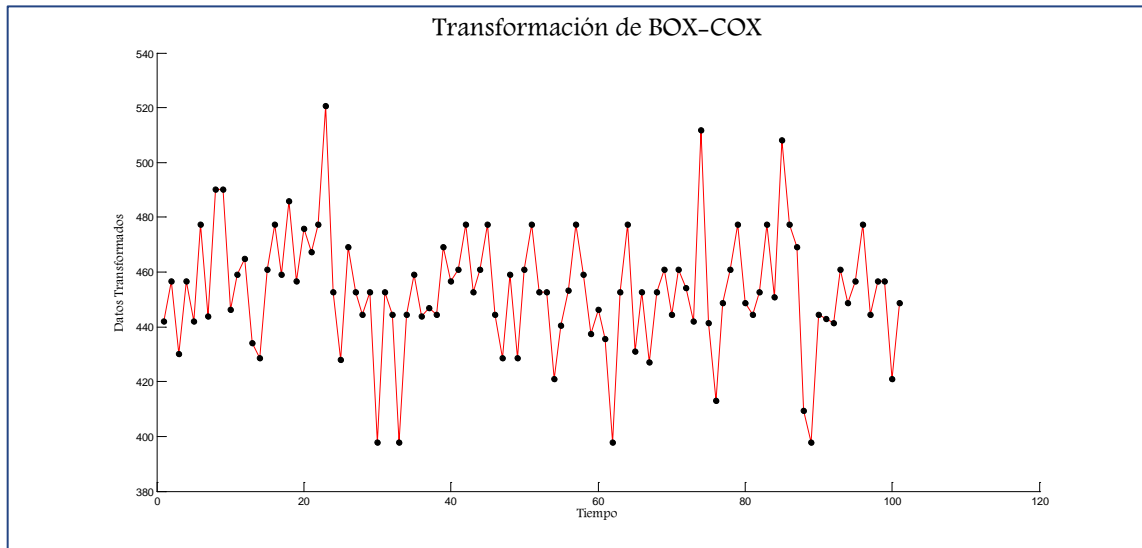


Figura 5.23. Representación gráfica de los datos transformados.

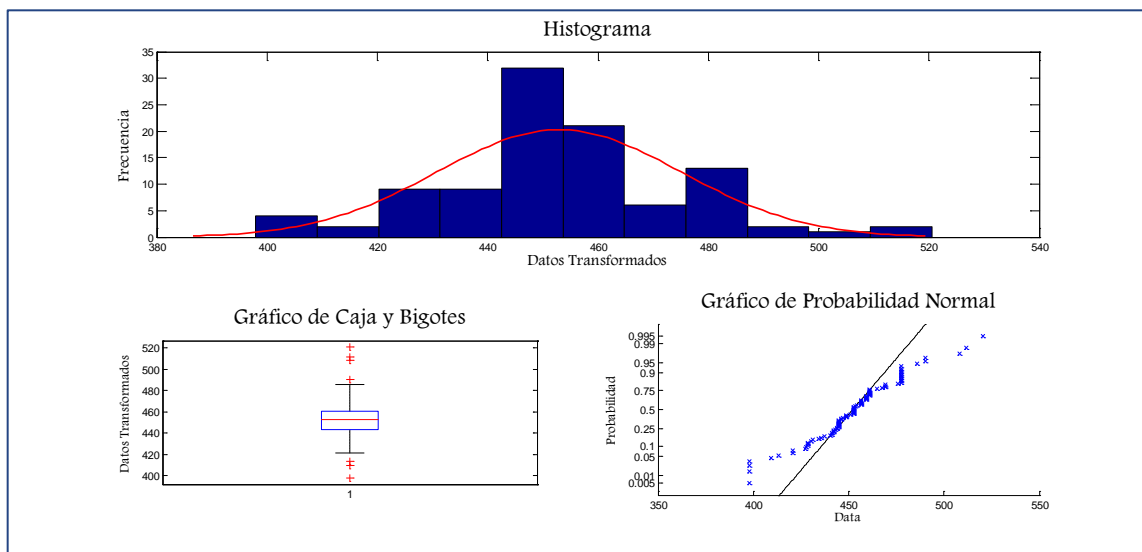


Figura 5.24. Representación gráfica de los datos transformados.

Los datos no son estacionarios en su forma original.



¿Autoriza la realización de una 1ª diferenciación?:  
Si(1) No(0) 1  
¿Desea observar los datos diferenciados 1ª vez?: Si(1)  
No(0) 1

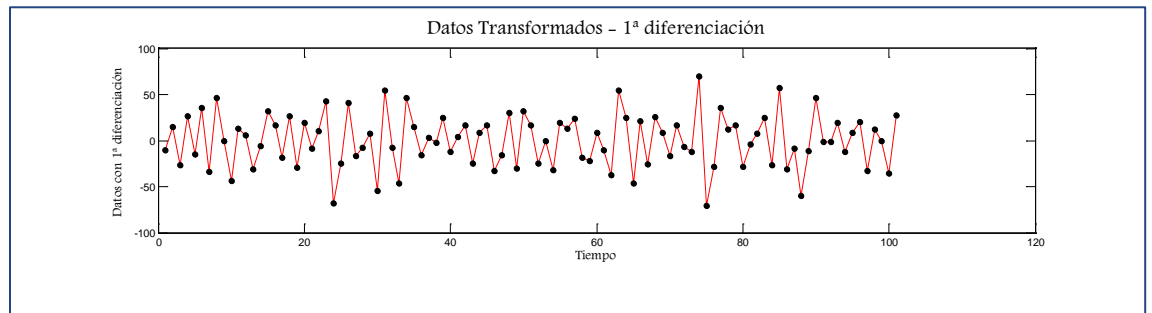


Figura 5.25. Representación gráfica de los datos diferenciados.

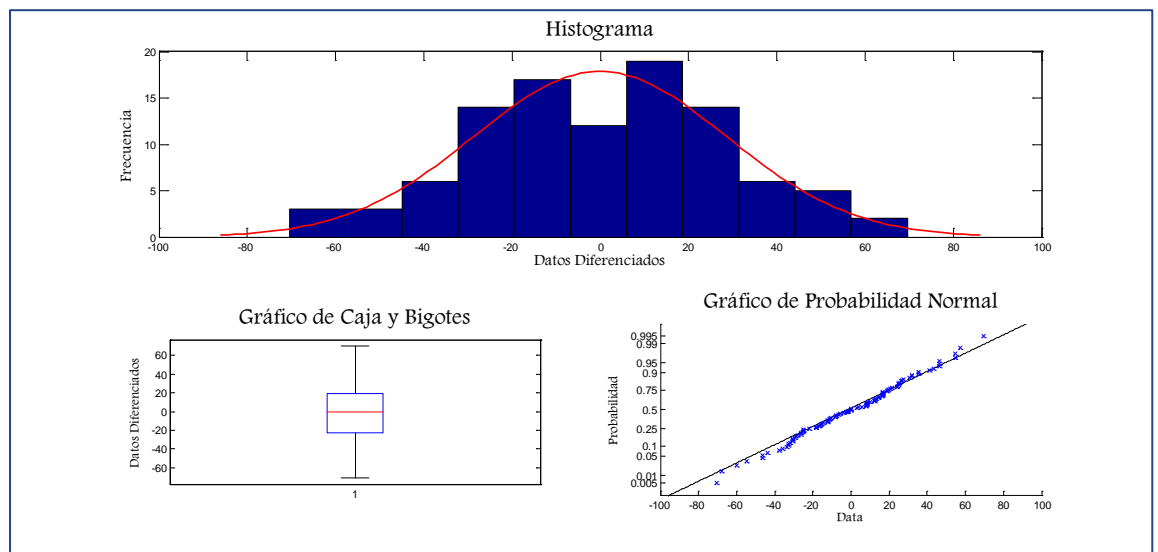


Figura 5.26. Representación gráfica de los datos diferenciados.

Los datos son estacionarios en 1ª diferenciación



A continuación se aplicará la prueba h-DURBIN para determinar si los datos son o no autocorrelacionados.

¿Autoriza la realización de esta prueba?: Si(1) No(0) 1

No se puede aplicar la prueba h-DURBIN porque el denominador del estadístico h es menor a 0.

Se puede realizar la prueba de DURBIN-WATSON sencilla  
¿Autoriza la aplicación de dicha prueba?: Si(1) No(0) 1

A continuación se leerá el archivo que contiene los valores de du que son límites de la prueba, para su aplicación.

Los datos presentan autocorrelación negativa

¿Desea ignorar el fenómeno de autocorrelacion?: Si(1)  
No(0) 0

ignauto =

0

**EJEMPLO N°4. Tiempo de espera en un centro telefónico de atención al cliente.**





En este ejemplo, se busca mostrar una serie de tiempo con distribución Normal, donde sea de utilidad el hecho de contar con el uso en conjunto del APC con el gráfico de control CUSUM. También se busca aplicar la transformación de Box-Cox para eliminar la no estacionariedad en la varianza y la primera diferenciación para eliminar la no estacionariedad en la media.

La tabla 5.4 contiene los valores del tiempo de espera en un centro telefónico de atención al cliente. La data suministrada indica el tiempo en segundos, desde que se realiza la conexión telefónica (desde que contestan la llamada) hasta que se es atendido por el personal técnico (incluyendo tiempo de espera con la típica música de fondo).

Los datos fueron tomados en orden cronológico cada 20 min desde las 8am hasta la 1pm. La muestra fue tomada de diferentes operarios de la estación en forma aleatoria. Por tratarse de operarios, llega un momento en que estos se agotan y por esta razón se sospecha que el tiempo de espera puede aumentar en las últimas horas de la jornada laboral diaria. A continuación se muestra el conjunto de datos obtenido:

**Tabla 5.4.** Datos de tiempo de espera de llamadas telefónicas.

N°	Tiempo (seg)	N°	Tiempo (seg)	N°	Tiempo (seg)
1	165	11	252	21	177
2	255	12	270	22	135
3	266	13	216	23	181



4	270	14	180	24	200
5	185	15	192	25	189
6	195	16	225	26	150
7	201	17	131	27	103
8	250	18	160	28	125
9	182	19	241	29	290
10	175	20	195	30	240

Fuente: Mason y Young (2002).

Al ejecutar el programa, aparece en pantalla la primera instrucción:

A continuación debe leer el archivo que contiene los datos a estudiar para leerlos debe colocar el nombre del archivo EXCEL que contiene los datos a estudiar, hágalo cuando se le indique

Indique el nombre del archivo a leer: Ejemplo1.xlsx

Los datos son aleatorios

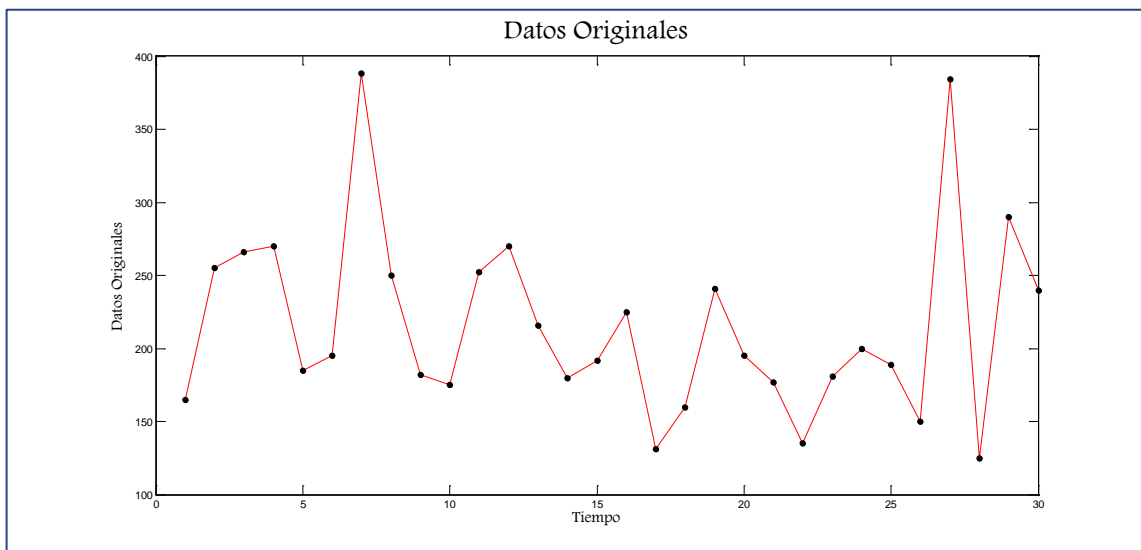


Figura 5.27. Representación gráfica de los datos en su forma original.

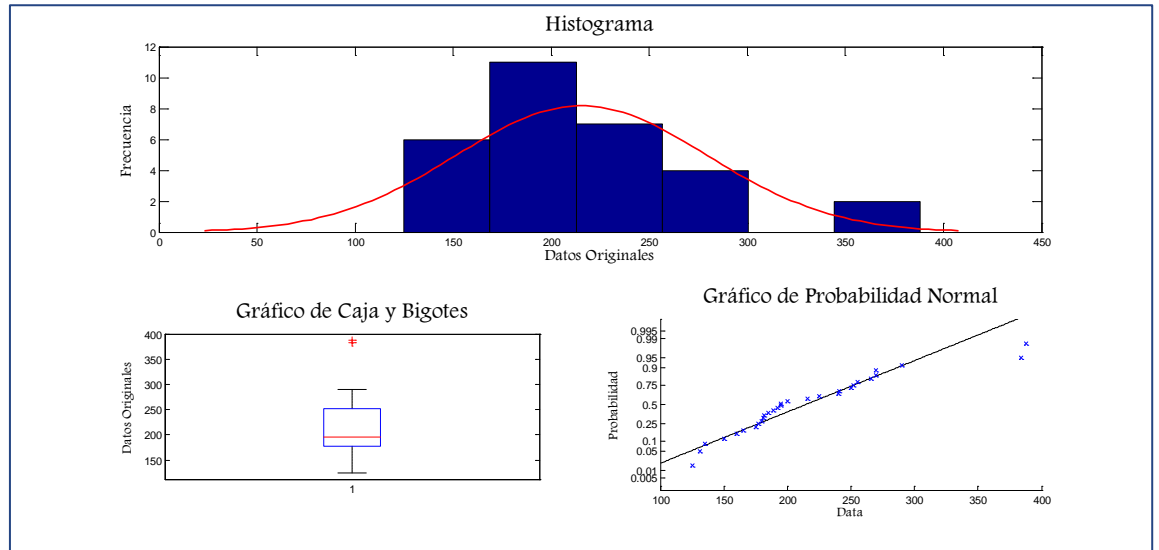


Figura 5.28. Representación gráfica de los datos en su forma original.

A continuación se realizará la eliminación de Outliers del vector original.

¿Autoriza la aplicación de dicha transformación?:  
Si(1) No(0) 1

Indique el valor de alpha: 0.05

Indique el valor del número de outliers que espera encontrar: 7

outliers detectados=

2

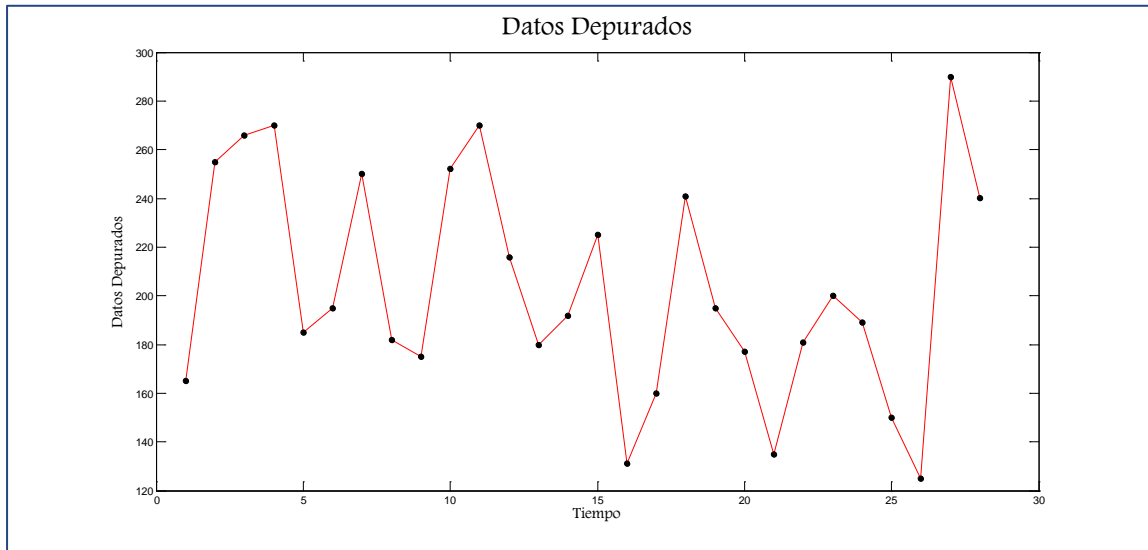


Figura 5.29. Representación gráfica de los datos depurados.

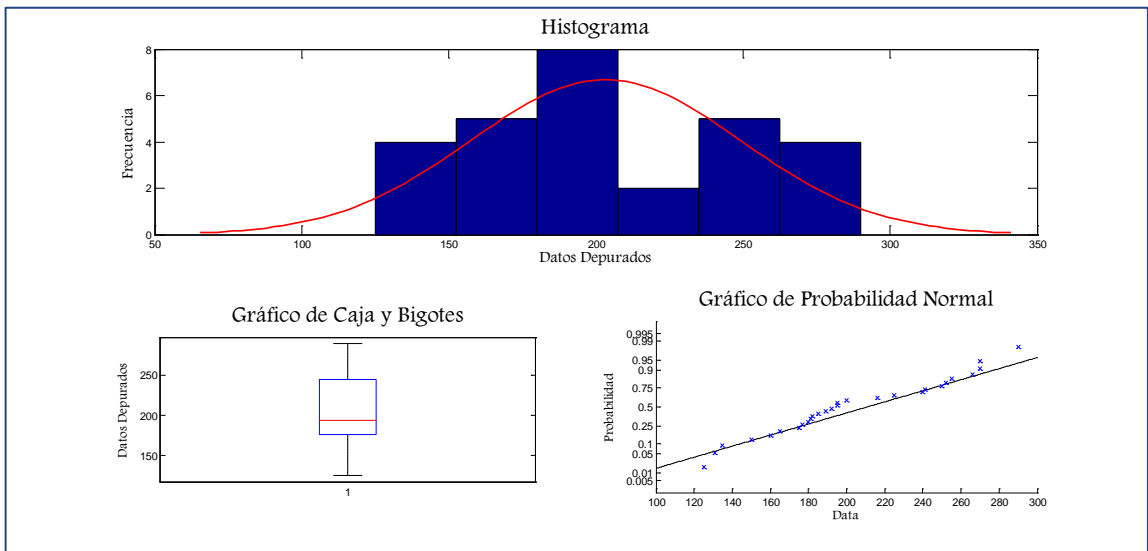


Figura 5.30. Representación gráfica de los datos depurados.

Al aplicar la prueba estadística de Lilliefors con un nivel de



confianza del 95% (Alfa=5%), no hay suficiente evidencia estadística para rechazar la hipótesis que los datos proceden de una Distribución Normal; dado que el p-valor del test realizado es mayor a 0.05

El valor de p es: 0.14392

Advertencia: Se recomienda el uso del gráfico EWMA por ser robusto en la presencia de distribuciones no Normales, a diferencia del CUSUM y Análisis de punto de cambio.

A continuación se realizará la transformación de BOX-COX que elimina la no estacionariedad en la varianza.

¿Autoriza la aplicación de dicha transformación?:  
Si(1) No(0) 1

¿Desea graficar la serie estacionaria en cuanto a la Varianza?: Si(1) No(0) 1

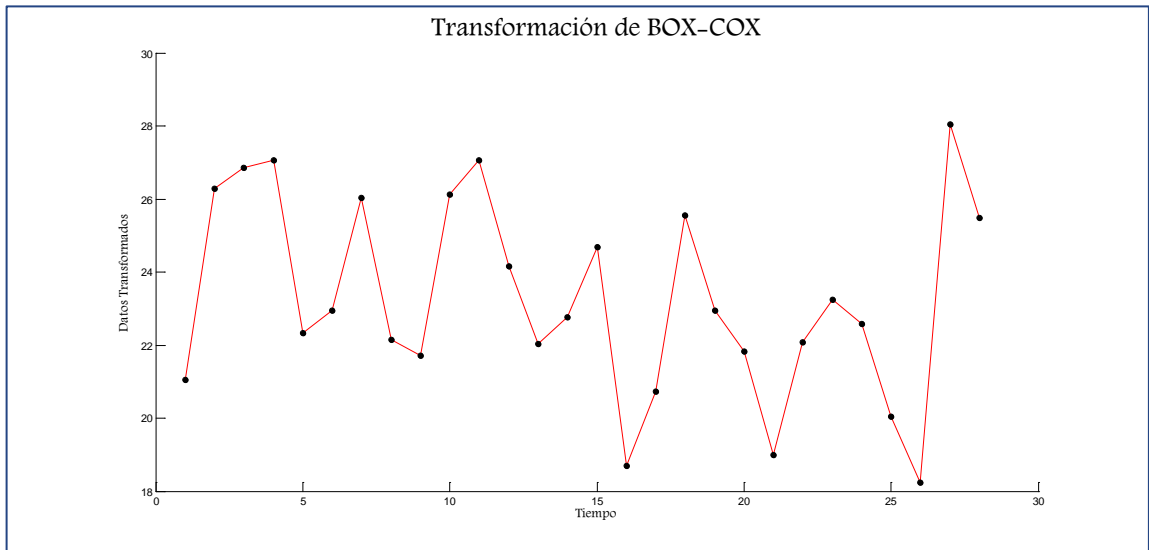


Figura 5.31. Representación gráfica de los datos transformados.

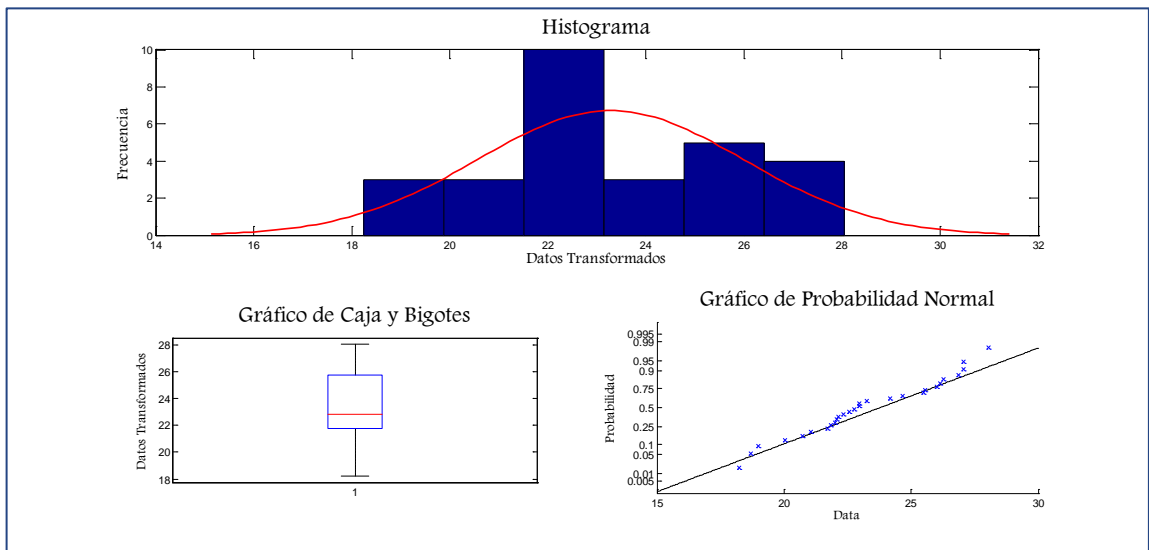


Figura 5.32. Representación gráfica de los datos transformados.

Los datos no son estacionarios en su forma original.  
¿Autoriza la realización de una 1ª diferenciación?:  
Si(1) No(0) 1



¿Desea observar los datos diferenciados 1ª vez?: Si(1)  
No(0) 1

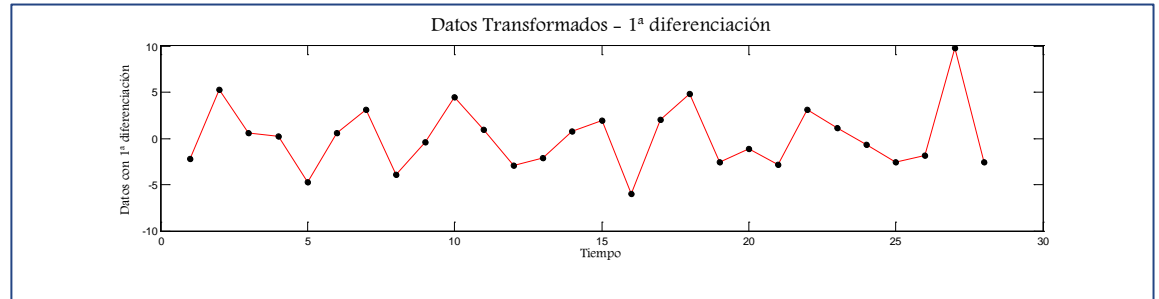


Figura 5.33. Representación gráfica de los datos diferenciados.

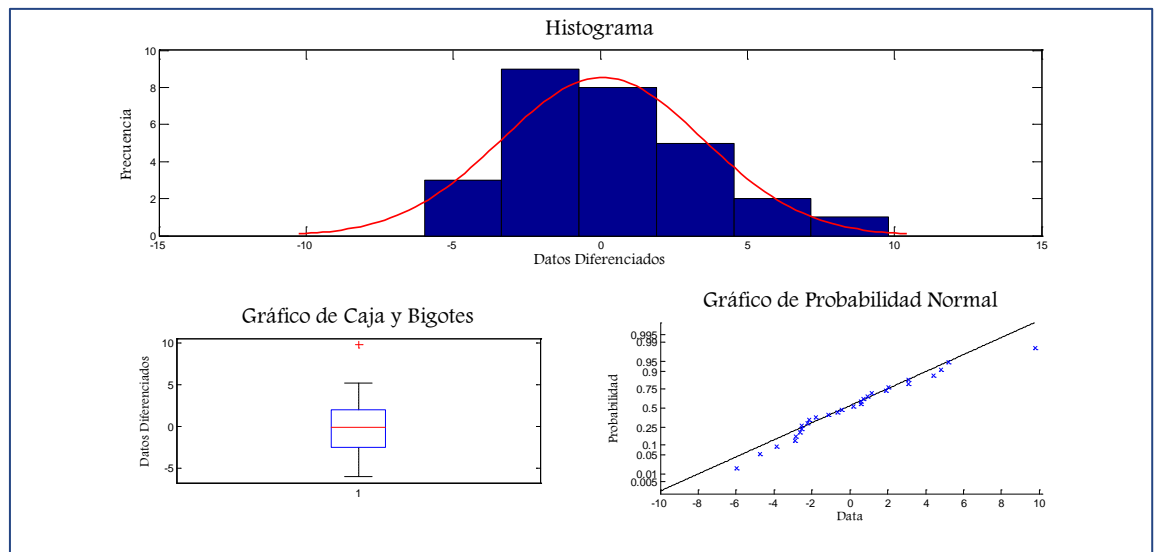


Figura 5.34. Representación gráfica de los datos diferenciados.

Los datos son estacionarios en 1ª diferenciación

A continuación se aplicará la prueba h-DURBIN para determinar si los datos son o no autocorrelacionados.



¿Autoriza la realización de esta prueba?: Si(1) No(0)  
1

No se puede aplicar la prueba h-DURBIN porque el denominador del estadístico h es menor a 0.

Se puede realizar la prueba de DURBIN-WATSON sencilla  
¿Autoriza la aplicación de dicha prueba?: Si(1) No(0)  
1

A continuación se leerá el archivo que contiene los valores de  $du$  que son límites de la prueba, para su aplicación.

Los datos no son autocorrelacionados

A continuación se diseñarán los Gráficos de Control. Para el Diseño del Gráfico EWMA con el Análisis de Punto de Cambio, seleccione (1); y para el Diseño del Gráfico CUSUM con el Análisis de Punto de Cambio, seleccione (2): 1

A continuación se mostrará la media de los datos:  
203.2857

D Significa la magnitud del cambio en la media que debe ser detectada lo más rápido posible.





Valores de D permitidos

0 0.25 0.5 0.75 1 1.25 1.5 1.75 2 2.25 2.5 2.75 3 3.25  
3.5 3.75 4

Indique el valor de D: 2

A continuación se leerá la tabla que contiene los valores de Lamda y K.

¿Desea ver el gráfico EWMA? Si(1) No(0) 1

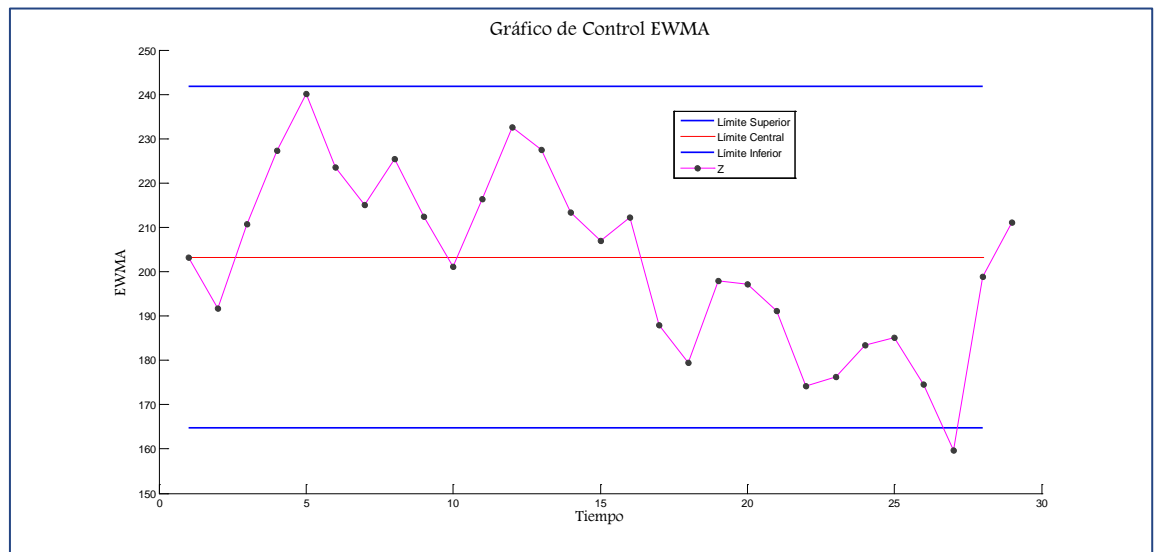


Figura 5.35. Representación del gráfico de control EWMA.

A continuación se procederá a construir el Gráfico de Control de  
Análisis de Punto de Cambio.

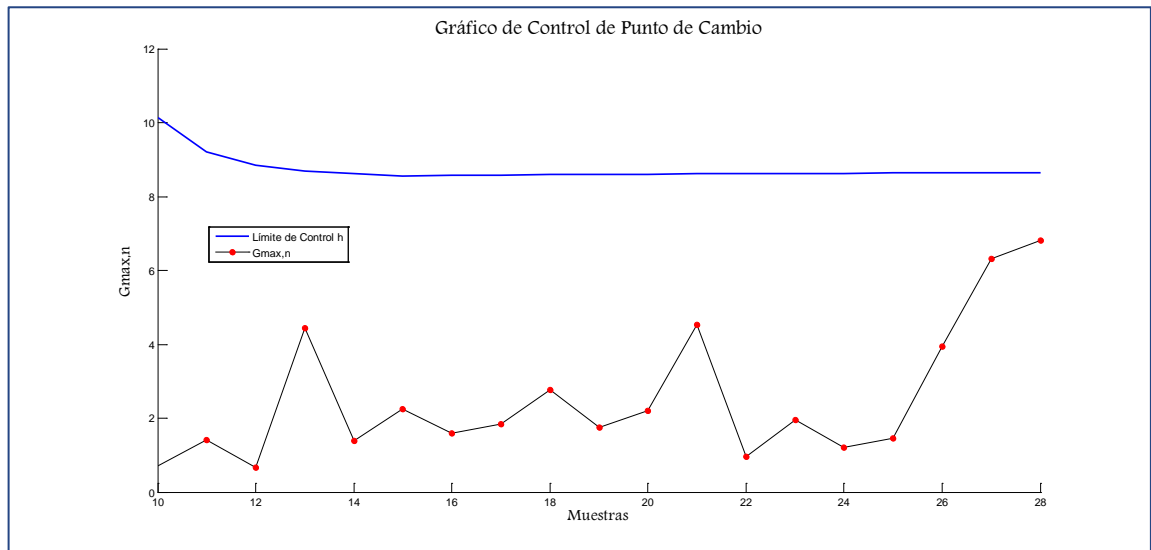


Deberá indicar el valor del alfa (probabilidad especificada de emitir una señal errónea)

Valores permitidos de alfa:

0.001 0.002 0.005 0.01 0.02 0.05

Indique el valor de alfa: 0.05



**Figura 5.36.** Representación del gráfico de control Análisis de Punto de Cambio.

No existe evidencia estadística para afirmar que ocurrió un cambio en la media o la varianza

El proceso se encuentra bajo control

La Media para los datos es: 203.2857

La Varianza para los datos es: 2110.3598



A continuación se muestran las tablas bibliográficas consultadas en el desarrollo de la herramienta computacional.

**TABLA A.1.** Correspondiente al estadístico  $d$  de Durbin Watson: puntos significativos de DU al nivel de significancia de 0.05.

K	1	2	3	4	5	6	7	8	9	10
N	DU	DU	DU	DU	DU	DU	DU	DU	DU	DU
26	1,222	1,312	1,411	1,518	1,635	1,759	1,889	2,026	2,168	2,313
30	1,489	1,567	1,65	1,739	1,833	1,931	2,034	2,141	2,251	2,363
31	1,449	1,57	1,65	1,735	1,825	1,92	2,018	2,12	2,226	2,333
32	1,502	1,574	1,65	1,732	1,819	1,909	2,004	2,102	2,203	2,306
33	1,508	1,577	1,651	1,73	1,813	1,9	1,991	2,085	2,181	2,281
34	1,514	1,58	1,652	1,728	1,808	1,891	1,979	2,069	2,162	2,257
35	1,519	1,584	1,653	1,726	1,803	1,884	1,967	2,054	2,144	2,236
36	1,525	1,587	1,654	1,724	1,799	1,877	1,957	2,041	2,127	2,216
37	1,53	1,59	1,655	1,723	1,795	1,87	1,948	2,029	2,221	2,198
38	1,535	1,594	1,656	1,722	1,792	1,864	1,939	2,017	2,098	2,18
39	1,54	1,597	1,658	1,722	1,789	1,859	1,932	2,007	2,085	2,164
40	1,544	1,6	1,659	1,721	1,786	1,854	1,924	1,997	2,072	2,149
45	1,566	1,615	1,666	1,72	1,776	1,835	1,895	1,958	2,022	2,088
50	1,585	1,628	1,674	1,721	1,771	1,822	1,875	1,93	1,986	2,044
55	1,601	1,641	1,681	1,724	1,768	1,814	1,861	1,909	1,959	2,01
60	1,616	1,652	1,689	1,727	1,767	1,808	1,85	1,894	1,939	1,984

**Fuente:** DAMODAR GUJARATI (2004).



**TABLA A.2.** Correspondiente a un ARL de 370 para la elección de los valores. de  $L$  y  $K$  en la Construcción del grafico EWMA para datos independientes



	$\lambda$	0,05	0,1	0,15	0,2	0,25	0,3	0,35	0,4	0,45	0,5
<b>D</b>	K	2,47	2,701	2,8	2,859	2,898	2,925	2,944	2,959	2,969	2,978
<b>0</b>	400	370	370	370	370	370	370	370	370	370	370
<b>0,25</b>	400	71,3	89,2	105,5	121	135	149,1	161,6	173,9	184,9	196,2
<b>0,5</b>	400	26	28,2	31,8	36,2	41,1	46,5	52,3	58,5	64,8	71,7
<b>0,75</b>	400	15,1	14,7	15,3	16,4	18	19,8	22	24,4	27,1	30,2
<b>1</b>	400	10,6	9,7	9,6	9,8	10,2	10,9	11,7	12,7	13,9	15,2
<b>1,25</b>	400	8,2	7,3	6,9	6,8	6,9	7,1	7,4	7,8	8,4	9
<b>1,5</b>	400	6,7	5,8	5,4	5,2	5,2	5,2	5,3	5,5	5,7	6
<b>1,75</b>	400	5,7	4,8	4,5	4,2	4,1	4,1	4,1	4,2	4,2	4,4
<b>2</b>	400	4,9	4,2	3,8	3,6	3,5	3,4	3,4	3,3	3,4	3,4
<b>2,25</b>	400	4,4	3,7	3,3	3,1	3	2,9	2,8	2,8	2,8	2,8
<b>2,5</b>	400	3,9	3,3	3	2,8	2,6	2,6	2,5	2,4	2,4	2,4
<b>2,75</b>	400	3,6	3	2,7	2,5	2,4	2,3	2,2	2,2	2,1	2,1
<b>3</b>	400	3,3	2,8	2,5	2,3	2,2	2,1	2	1,9	1,9	1,9
<b>3,25</b>	400	3,1	2,6	2,3	2,1	2	1,9	1,8	1,8	1,7	1,7
<b>3,5</b>	400	2,9	2,4	2,2	2	1,9	1,8	1,7	1,6	1,6	1,5
<b>3,7</b>	400	2,7	2,2	2,1	1,9	1,8	1,7	1,6	1,5	1,4	1,4



<b>5</b>	0										
<b>4</b>	40 0	2,5	2,1	2	1,8	1,7	1,6	1,5	1,4	1,3	1,3

Fuente: CROWDER, S. V.; 1987.

**TABLA A.2. (Continuación)** Correspondiente a un ARL de 370 para la elección de los valores de L y K en la construcción del grafico EWMA para datos independientes.

	$\lambda$	0,55	0,6	0,65	0,7	0,75	0,8	0,85	0,9	0,95	1
<b>D</b>	K	2,984	2,988	2,992	2,994	2,996	2,998	2,999	2,999	3	3
<b>0</b>	400	370	370	370	370	370	370	370	370	370	370
<b>0,25</b>	400	206,4	215,9	225,6	234,2	242,8	251,5	259,5	266,5	274,3	281,2
<b>0,5</b>	400	78,7	86	93,7	101,5	109,8	118,5	127,3	136,2	145,7	155,2
<b>0,75</b>	400	33,6	37,2	41,3	45,7	50,5	55,8	61,4	67,4	74,1	81,2
<b>1</b>	400	16,8	18,6	20,6	22,9	25,4	28,4	31,6	35,2	39,4	43,9
<b>1,25</b>	400	9,7	10,6	11,6	12,8	14,2	15,8	17,6	19,7	22,2	25
<b>1,5</b>	400	6,3	6,8	7,3	7,9	8,7	9,6	10,6	11,8	13,3	15
<b>1,75</b>	400	4,5	4,8	5	5,4	5,8	6,3	6,9	7,6	8,4	9,5
<b>2</b>	400	3,5	3,6	3,7	3,9	4,1	4,4	4,8	5,2	5,7	6,3
<b>2,25</b>	400	2,8	2,9	2,9	3	3,1	3,3	3,5	3,7	4	4,4
<b>2,5</b>	400	2,4	2,4	2,4	2,5	2,5	2,6	2,7	2,8	3	3,2



<b>2,75</b>	400	2,1	2	2	2,1	2,1	2,1	2,2	2,3	2,4	2,5
<b>3</b>	400	1,8	1,8	1,8	1,8	1,8	1,8	1,8	1,9	1,9	2
<b>3,25</b>	400	1,6	1,6	1,6	1,6	1,6	1,6	1,6	1,6	1,6	1,7
<b>3,5</b>	400	1,5	1,5	1,4	1,4	1,4	1,4	1,4	1,4	1,4	1,4
<b>3,75</b>	400	1,4	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3	1,3
<b>4</b>	400	1,3	1,3	1,2	1,2	1,2	1,2	1,2	1,2	1,2	1,2

Fuente: CROWDER, S. V.; 1987.

**TABLA A.3.** Correspondiente a un  $\psi=0.5$  para la elección de los valores de  $\lambda$  en la construcción de grafico EWMA para datos autocorrelacionados

$\Psi$	$\Psi$	$\Psi$	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
$\bar{\delta}$	$\bar{\delta}$	$\bar{\delta}$	0,5	1	1,5	2	2,5	3	4	5
c	$\phi$	$\lambda$								
<b>3,085</b>	0,4	0,05	<b>36,98</b>	14,06	8,54	6,21	4,93	4,08	3,13	2,55
<b>3,289</b>	0,4	0,1	40,48	<b>13,27</b>	7,51	5,18	4,03	3,33	2,54	2,08



<b>3,375</b>	0,4	0,15	48,17	13,52	7,11	4,75	3,63	2,98	2,24	1,85
<b>3,391</b>	0,4	0,2	53,68	14,09	<b>6,91</b>	4,49	3,35	2,7	2,02	1,67
<b>3,394</b>	0,4	0,25	62,3	14,95	6,97	4,35	3,21	2,57	1,9	1,55
<b>3,387</b>	0,4	0,3	68,11	16,19	7,12	4,3	3,1	2,44	1,79	1,45
<b>3,375</b>	0,4	0,35	75,72	17,59	7,31	<b>4,24</b>	3,03	2,35	1,7	1,37
<b>3,355</b>	0,4	0,4	81,21	19,01	7,59	4,33	2,94	2,26	1,61	1,29
<b>3,299</b>	0,4	0,5	92,67	22,25	8,43	4,42	<b>2,9</b>	<b>2,16</b>	1,5	1,17
<b>3,07</b>	0,4	0,8	160,89	46,95	16,35	7,22	3,64	2,2	<b>1,22</b>	<b>1,03</b>

Fuente: LU, CHAO- WEN and REYNOLDS, JR. M. R. ,1999.

**TABLA A.3(Continuación).** Correspondiente a un  $\psi=0.5$  para la elección de los valores de  $\lambda$  en la construcción de grafico EWMA para datos autocorrelacionados





$\Psi$	$\psi$	$\psi$	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
$\Delta$	$\delta$	$\delta$	0,5	1	1,5	2	2,5	3	4	5
C	$\varphi$	$\lambda$								
<b>4,591</b>	0,8	0,5	<b>78,13</b>	<b>27,06</b>	14,88	10,09	7,72	6,23	4,57	3,7
<b>4,625</b>	0,8	0,1	91,58	28,16	13,52	8,54	6,16	4,93	3,56	2,81
<b>4,536</b>	0,8	0,15	100,39	29,74	13,4	7,89	5,52	4,22	3,01	2,4
<b>4,375</b>	0,8	0,2	107,41	30,77	<b>13,18</b>	7,39	5,02	3,8	2,64	2,11
<b>4,272</b>	0,8	0,25	115,22	33,49	13,93	7,37	4,75	3,55	2,43	1,92
<b>4,143</b>	0,8	0,3	121,47	34,75	14,25	7,33	4,59	3,3	2,23	1,76
<b>4,015</b>	0,8	0,35	127,23	36,24	14,2	<b>7,31</b>	4,43	3,14	2,06	1,62
<b>3,906</b>	0,8	0,4	132,79	37,71	14,75	7,34	4,3	3	1,94	1,51
<b>3,707</b>	0,8	0,5	139,36	40,45	15,49	7,43	<b>4,22</b>	2,84	1,75	1,34
<b>3,35</b>	0,8	0,8	176,27	56,35	21,81	9,49	4,8	<b>2,59</b>	<b>1,24</b>	<b>1,03</b>

Fuente: LU, CHAO- WEN and REYNOLDS, JR. M. R. ,1999.



**TABLA A.3 (Continuación).** Correspondiente a un  $\psi=0.9$  para la elección de los valores de  $\lambda$  en la construcción de grafico EWMA para datos autocorrelacionados

$\Psi$	$\psi$	$\psi$	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9
$\Delta$	$\delta$	$\delta$	0,5	1	1,5	2	2,5	3	4	5
C	$\phi$	$\lambda$								
<b>3,479</b>	0,4	0,05	<b>44,9</b>	16,48	9,96	7,16	5,54	4,63	3,46	2,84
<b>3,684</b>	0,4	0,1	51,63	<b>16,16</b>	8,83	6	4,6	3,76	2,79	2,29
<b>3,748</b>	0,4	0,15	60,83	16,92	8,46	5,58	4,13	3,35	2,46	2,03
<b>3,75</b>	0,4	0,2	68,42	18,04	<b>8,4</b>	5,28	3,88	3,07	2,25	1,83
<b>3,732</b>	0,4	0,25	76,88	19,34	8,56	5,18	3,68	2,88	2,08	1,69
<b>3,703</b>	0,4	0,3	84,97	20,92	8,8	5,14	3,59	2,75	1,97	1,58
<b>3,656</b>	0,4	0,35	89,75	22,31	9,23	<b>5,06</b>	3,47	2,64	1,84	1,48
<b>3,622</b>	0,4	0,4	99,04	24,1	9,56	5,23	3,46	2,56	1,78	1,4
<b>3,513</b>	0,4	0,5	110,42	27,42	10,54	5,42	<b>3,39</b>	2,43	1,62	1,26
<b>3,21</b>	0,4	0,8	162,47	48,17	17,46	7,75	4	<b>2,31</b>	<b>1,23</b>	<b>1,03</b>

**Fuente:** LU, CHAO- WEN and REYNOLDS, JR. M. R. ,1999.



**TABLA A.3 (Continuación).** Correspondiente a un  $\psi=0.9$  para la elección de los valores de  $\lambda$  y C en la construcción de gráfico EWMA para datos autocorrelacionados

$\psi$	$\psi$	$\psi$	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9
$\delta$	$\delta$	$\delta$	0,5	1	1,5	2	2,5	3	4	5
c	$\varphi$	$\lambda$								
<b>5,65</b> 4	0,8	0,5	<b>108,4</b>	<b>38,58</b>	20,5	13,55	9,95	7,92	5,69	4,5
<b>5,64</b> 1	0,8	0,1	127,6	41,76	<b>20,11</b>	12,06	8,3	6,44	4,43	3,45
<b>5,43</b>	0,8	0,15	139,14	45,17	20,64	11,5	7,64	5,59	3,73	2,88
<b>5,20</b> 3	0,8	0,2	148,3	48,13	21,54	<b>11,41</b>	7,25	5,12	3,29	2,52
<b>4,98</b> 1	0,8	0,2 5	154,44	50,45	21,48	11,48	6,87	4,77	2,96	2,25
<b>4,78</b> 1	0,8	0,3 2	160,0	52,81	22,68	11,55	6,69	4,48	2,72	2,05
<b>4,57</b>	0,8	0,3 5	161,74	54,82	22,76	11,43	6,58	4,27	2,49	1,87
<b>4,39</b> 4	0,8	0,4	168,85	55,86	23,12	11,62	6,52	4,06	2,3	1,72
<b>4,07</b> 4	0,8	0,5	174,01	58,58	24,27	11,66	6,32	3,95	2,05	1,49
<b>4,2</b>	0,8	0,8	187,08	65,77	26,74	12,26	<b>5,95</b>	<b>3,2</b>	<b>1,31</b>	<b>1,02</b>

Fuente: LU, CHAO- WEN and REYNOLDS, JR. M. R. ,1999.



**TABLA A.4.** Correspondiente a la obtención del valor de  $c$  en el gráfico EWMA Autocorrelacionado

<b>lamda</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>	<b>0,1</b>
<b><math>\phi/\psi</math></b>	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
<b>0,1</b>	2,6	2,8	2,9	3	2,9	2,9	2,9	2,9	2,9
<b>0,2</b>	2,7	2,85	2,95	3,1	3	3,1	3,2	3,3	3,3
<b>0,3</b>	2,75	2,9	3	3,2	3,2	3,3	3,4	3,4	3,4
<b>0,4</b>	2,8	3	3,1	3,3	3,4	3,4	3,5	3,6	3,7
<b>0,5</b>	2,9	3,1	3,2	3,4	3,5	3,6	3,8	3,9	4
<b>0,6</b>	2,95	3,2	3,4	3,6	3,7	4	4,2	4,4	4,5
<b>0,7</b>	3	3,3	3,6	3,9	4,3	4,4	4,6	4,8	5
<b>0,8</b>	3,2	3,5	4	4,1	4,6	4,9	5,2	5,4	5,7



<b>0,9</b>	3,5	4	4,5	5	5,3	5,6	6	6,3	6,5
<b>lamda</b>	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2	0,2
<b>φ/ψ</b>	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
<b>0,1</b>	2,9	3	3	3,1	3,1	3,2	3,3	3,3	3,4
<b>0,2</b>	2,85	3,1	3,1	3,2	3,2	3,3	3,4	3,4	3,45
<b>0,3</b>	2,9	3,2	3,25	3,3	3,3	3,35	3,45	3,5	3,55
<b>0,4</b>	3	3,25	3,3	3,35	3,4	3,5	3,6	3,7	3,9
<b>0,5</b>	3,1	3,3	3,4	3,5	3,6	3,7	3,8	3,9	4,3
<b>0,6</b>	3,2	3,35	3,5	3,6	3,8	3,9	4,2	4,3	4,4
<b>0,7</b>	3,3	3,4	3,7	3,9	4,2	4,3	4,4	4,6	4,7
<b>0,8</b>	3,4	3,6	3,9	4,3	4,4	4,7	4,9	5,1	5,3
<b>0,9</b>	3,5	3,9	4,2	4,5	4,8	5,1	5,3	5,4	5,6

Fuente: LU, CHAO- WEN and REYNOLDS, JR. M. R. ,1999.

**TABLA A.4. (Continuación).** Correspondiente a la obtención del valor de c en el gráfico EWMA Autocorrelacionado

<b>lamda</b>	<b>0,3</b>	<b>0,3</b>	<b>0,3</b>	<b>0,3</b>	<b>0,3</b>	<b>0,3</b>	<b>0,3</b>	<b>0,3</b>	<b>0,3</b>
<b>φ/ψ</b>	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
<b>0,1</b>	3	3,15	3,2	3,2	3,25	3,25	3,3	3,35	3,35



<b>0,2</b>	3,1	3,15	3,25	3,3	3,35	3,35	3,35	3,4	3,4
<b>0,3</b>	3,15	3,2	3,3	3,35	3,4	3,4	3,4	3,5	3,55
<b>0,4</b>	3,2	3,25	3,35	3,4	3,45	3,45	3,5	3,6	3,7
<b>0,5</b>	3,25	3,3	3,45	3,5	3,6	3,6	3,7	3,8	3,9
<b>0,6</b>	3,3	3,35	3,5	3,6	3,7	3,8	3,9	4,1	4,2
<b>0,7</b>	3,35	3,4	3,6	3,7	4,2	4,2	4,3	4,4	4,5
<b>0,8</b>	3,4	3,5	3,7	3,9	4,2	4,4	4,5	4,6	4,8
<b>0,9</b>	3,45	3,6	3,9	4,2	4,4	4,6	4,7	4,8	5

**Fuente:** LU, CHAO- WEN and REYNOLDS, JR. M. R. ,1999.

**TABLA A.5.** Correspondiente a la elección de los valores de  $r$  en gráficos de CUSUM autocorrelacionado.



$\Phi$	$\Phi$	0,4	0,4	0,4	0,4	0,4	0,4	0,4	0,4
$\Psi$	$\Psi$	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
$\Delta$	r	0,5	1	1,5	2	2,5	3	4	5
<b>18,72</b>	0,1	38,35	17,73	11,35	8,5	6,82	5,7	4,35	3,55
<b>13,15</b>	0,2	36,05	14,68	9,21	6,76	5,35	4,5	3,41	2,77
<b>10,05</b>	0,3	37,23	13,54	8,06	5,82	4,58	3,79	2,87	2,36
<b>8,05</b>	0,4	40,33	12,9	7,38	5,21	4,04	3,34	2,53	2,06
<b>6,71</b>	0,5	45,61	13	7,12	4,79	3,68	3,02	2,27	1,9
<b>5,67</b>	0,6	50,9	13,45	6,81	4,57	3,43	2,8	2,1	1,78
<b>4,88</b>	0,7	58,11	14,12	6,75	4,35	3,25	2,62	1,96	1,61
<b>4,22</b>	0,8	64,49	14,98	6,82	4,18	3,05	2,46	1,83	1,45
<b>3,69</b>	0,9	71,46	15,94	6,87	4,12	2,97	2,33	1,7	1,3
<b>3,24</b>	1	77,63	17,13	7,1	4,07	2,87	2,24	1,59	1,21
<b>2,39</b>	1,25	93,88	21,2	7,78	4,1	2,73	2,04	1,37	1,09
<b>1,8</b>	1,5	109,04	26,03	8,96	4,39	2,68	1,91	1,26	1,05
$\Phi$	$\Phi$	0,4	0,4	0,4	0,4	0,4	0,4	0,4	0,4
$\Psi$	$\Psi$	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9
$\delta$	r	0,5	1	1,5	2	2,5	3	4	5
<b>22,33</b>	0,1	46,12	21,1	13,69	10,05	8,14	6,71	5,11	4,15
<b>15,86</b>	0,2	43,14	17,71	11,03	8,01	6,33	5,27	3,98	3,23
<b>12,28</b>	0,3	44,65	16,26	9,71	6,98	5,48	4,49	3,4	2,76
<b>9,97</b>	0,4	49,15	15,95	9,16	6,38	4,87	4,01	3	2,44
<b>8,26</b>	0,5	54,29	15,95	8,58	5,88	4,45	3,62	2,7	2,17
<b>6,7</b>	0,6	60,1	16,39	8,4	5,53	4,14	3,33	2,45	2,01
<b>6,01</b>	0,7	66,86	17,29	8,23	5,3	3,88	3,12	2,27	1,88
<b>5,2</b>	0,8	74,06	18,21	8,2	5,18	3,71	2,92	2,13	1,77
<b>4,53</b>	0,9	80,63	19,33	8,43	5,04	3,53	2,76	2,01	1,61
<b>3,96</b>	1	86,5	20,59	8,6	4,97	3,42	2,63	1,87	1,46
<b>2,87</b>	1,25	101,1	25,1	9,37	4,95	3,22	2,37	1,6	1,19
<b>2,1</b>	1,5	117,76	29,8	10,62	5,26	3,15	2,22	1,39	1,08

Fuente: LU, CHAO- WEN and REYNOLDS, JR. M. R. ,2001.



**TABLA A.5 (Continuación).** Correspondiente a la elección de los valores de  $r$  en gráficos de CUSUM autocorrelacionado.

$\Phi$	$\Psi$	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8
$\Psi$	$\Psi$	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5
$\delta$	r	0,5	1	1,5	2	2,5	3	4	5
<b>35,11</b>	0,1	74,71	34,6	22,03	16,26	12,88	10,72	8,01	6,48
<b>25,72</b>	0,2	71,31	29,44	18,17	13	10,21	8,36	6,23	5,03
<b>20,13</b>	0,3	74,21	27,58	16,04	11,32	8,76	7,13	5,26	4,22
<b>16,18</b>	0,4	77,79	26,83	14,93	10,12	7,78	6,25	4,57	3,67
<b>13,15</b>	0,5	83,3	26,77	14,16	8,27	6,94	5,51	4	3,19
<b>10,89</b>	0,6	89,44	26,94	13,5	8,6	6,29	5	3,57	2,84
<b>9,1</b>	0,7	94,94	28,23	13,22	8,08	5,82	4,54	3,19	2,54
<b>7,63</b>	0,8	101,7	28,52	13,05	7,67	5,32	4,12	2,88	2,25
<b>6,43</b>	0,9	105,94	29,97	13,21	7,36	4,98	3,79	2,61	2,05
<b>5,39</b>	1	113,48	31,08	13,2	7,07	4,64	3,48	2,35	1,9
<b>3,57</b>	1,25	124,78	35,45	13,6	6,84	4,12	2,89	1,91	1,42
<b>2,37</b>	1,5	135,28	38,63	14,67	6,77	3,88	2,56	1,51	1,13
$\Phi$	$\Psi$	0,8	0,8	0,8	0,8	0,8	0,8	0,8	0,8
$\Psi$	$\Psi$	0,9	0,9	0,9	0,9	0,9	0,9	0,9	0,9
$\delta$	r	0,5	1	1,5	2	2,5	3	4	5
<b>47,42</b>	0,1	101,13	47,54	30,09	22,12	17,46	14,45	10,76	8,66
<b>36,14</b>	0,2	96,28	41,3	25,69	18,36	14,18	11,7	8,6	6,92
<b>28,95</b>	0,3	99,65	39,1	23,09	16,17	12,41	10,05	7,37	5,83
<b>23,72</b>	0,4	103,44	37,96	21,75	14,57	11,16	8,91	6,39	5,08
<b>19,8</b>	0,5	109,76	38,65	20,62	13,64	10,13	8,07	5,77	4,55
<b>16,53</b>	0,6	116,28	38,82	20,17	12,93	9,33	7,3	5,15	4,04
<b>14,09</b>	0,7	122,86	40,4	20,05	12,22	8,75	6,76	4,68	3,63
<b>12,06</b>	0,8	128,54	42,35	20,03	11,88	8,29	6,26	4,28	3,3
<b>10,25</b>	0,9	134,12	43,32	19,97	11,57	7,76	5,73	3,86	2,99
<b>8,79</b>	1	141,65	45,21	20,22	11,24	7,42	5,39	3,55	2,72
<b>5,86</b>	1,25	152,25	49,48	20,98	10,9	6,68	4,58	2,79	2,14





<b>3,85</b>	1,5	162,18	53,93	21,9	11,02	6,17	4,02	2,31	1,68
-------------	-----	--------	-------	------	-------	------	------	------	------

Fuente: LU, CHAO- WEN and REYNOLDS, JR. M. R. ,2001.

**TABLA A.6.** Correspondiente a la elección de C en la obtención del gráfico CUSUM autocorrelacionados.

<b>r</b>	<b>0,25</b>	<b>0,25</b>	<b>0,25</b>	<b>0,25</b>	<b>0,25</b>	<b>0,25</b>	<b>0,25</b>	<b>0,25</b>	<b>0,25</b>
<b>φ/ψ</b>	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
<b>0,1</b>	7,7	7,8	8,2	8,3	8,6	8,7	8,8	8,9	9
<b>0,2</b>	7,7	7,8	8,4	8,7	9	9,3	9,5	9,8	10
<b>0,3</b>	7,7	8,5	8,7	9	10	10,2	10,7	11,1	11,5
<b>0,4</b>	7,7	8,8	9,1	9,8	10,7	11,3	12,4	13,1	14,3
<b>0,5</b>	7,7	9,1	10	10,5	11,5	12,9	14	15,2	16,2
<b>0,6</b>	7,7	9,9	10,9	12,5	13,5	15,1	16,1	17,4	19,2
<b>0,7</b>	7,7	10,2	12,5	14,5	15,3	17,7	19,5	21,1	22,5
<b>0,8</b>	7,7	11,2	14,9	17,5	20,01	22,5	25,03	27,9	30,01
<b>0,9</b>	7,7	14,7	19,7	24,7	28,2	33,3	37,1	41,1	44,9
<b>r</b>	<b>0,5</b>	<b>0,5</b>	<b>0,5</b>	<b>0,5</b>	<b>0,5</b>	<b>0,5</b>	<b>0,5</b>	<b>0,5</b>	<b>0,5</b>
<b>φ/ψ</b>	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
<b>0,1</b>	4,9	4,9	5,01	5,02	5,03	5,07	5,1	5,15	5,17
<b>0,2</b>	4,9	5,1	5,4	5,7	5,9	6,2	6,5	6,8	7,2
<b>0,3</b>	4,9	5,2	5,5	6,1	6,5	6,8	7,01	7,3	7,5
<b>0,4</b>	4,9	5,4	6,1	6,4	6,7	7,3	7,5	7,9	8,3
<b>0,5</b>	4,9	5,5	6,1	6,5	6,8	7,1	7,7	8,3	8,8
<b>0,6</b>	4,9	5,5	6,3	7,1	7,7	8,1	9,5	10,03	12,1
<b>0,7</b>	4,9	5,9	6,9	7,8	8,9	10,6	12,2	13,3	14,4
<b>0,8</b>	4,9	6,3	7,7	9,9	12,1	13,2	15,01	16,3	17,1
<b>0,9</b>	4,9	7,5	10,1	14	16,1	18,9	22,1	24,8	26,8



Fuente: LU, CHAO- WEN and REYNOLDS, JR. M. R. ,2001.

**TABLA A.6 (Continuación).** Correspondiente a la elección de C en la obtención del gráfico CUSUM autocorrelacionados.

$r$	1	1	1	1	1	1	1	1	1
$\phi/\psi$	0,1	0,2	0,3	0,4	0,5	0,6	0,7	0,8	0,9
0,1	2,6	2,61	2,62	2,63	2,7	2,8	2,83	2,85	2,9
0,2	2,6	2,65	2,7	2,75	2,8	2,85	2,9	2,98	3,2
0,3	2,6	2,7	2,8	2,9	3,1	3,2	3,3	3,45	3,6
0,4	2,6	2,7	2,8	2,9	3,1	3,2	3,6	3,8	4,2
0,5	2,6	2,88	3,16	3,44	3,72	4	4,29	4,57	4,9
0,6	2,6	2,9	3,1	3,3	3,8	4,6	4,8	5,1	6,01
0,7	2,6	2,9	0,32	3,5	3,95	4,8	5,1	5,8	6,3
0,8	2,6	3	3,8	4,2	4,8	5,4	6,3	7,1	8,1
0,9	2,6	3,2	3,9	4,8	5,9	7	8,3	9,8	10,1

Fuente: LU, CHAO- WEN and REYNOLDS, JR. M. R. ,2001.



**TABLA A.7.** Límite de Control  $hn, \alpha$  para una muestra de Tamaño  $n$  , rechazo  $\alpha$ , a partir de  $n = 10$ .

	<b><math>hn, \alpha</math></b>					
<b>n</b>	0,05	0,02	0,01	0,005	0,002	0,001
<b>10</b>	10,128	12,237	13,795	15,33	17,352	18,84
<b>11</b>	9,213	11,389	12,996	14,556	16,609	18,173
<b>12</b>	8,854	11,083	12,719	14,313	16,397	17,965
<b>13</b>	8,69	10,961	12,631	14,265	16,353	17,95
<b>14</b>	8,616	10,917	12,61	14,249	16,361	17,978
<b>15</b>	8,588	10,909	12,618	14,277	16,423	18,015
<b>16</b>	8,582	10,928	12,637	14,324	16,464	18,083
<b>17</b>	8,586	10,942	12,664	14,345	16,501	18,113
<b>18</b>	8,59	10,964	12,692	14,386	16,55	18,192
<b>19</b>	8,593	10,973	12,709	14,394	16,567	18,187
<b>20</b>	8,599	10,989	12,734	14,419	16,614	18,219
<b>22</b>	8,611	11,015	12,763	14,462	16,653	18,261
<b>24</b>	8,629	11,043	12,79	14,497	16,682	18,343



26	8,644	11,06	12,819	14,536	16,722	18,368
28	8,661	11,089	12,846	14,559	16,754	18,375
30	8,669	11,097	12,877	14,591	16,785	18,436
35	8,683	11,133	12,91	14,636	16,845	18,478
40	8,692	11,151	12,932	14,657	16,884	18,572
45	8,712	11,164	12,947	14,676	16,908	18,572
50	8,714	11,181	12,971	14,713	16,946	18,617
60	8,734	11,197	12,99	14,746	16,98	18,616
70	8,745	11,21	13,003	14,751	16,985	18,634
80	8,748	11,224	13,022	14,766	16,989	18,679
90	8,743	11,238	13,035	14,785	17,043	18,71
100	8,77	11,231	13,029	14,771	17,052	18,708
125		11,257	13,061	14,799	17,033	18,697
150		11,253	13,037	14,812	17,069	18,756
175		11,283	13,06	14,822	17,069	18,729
200		11,277	13,078	14,803	17,065	18,738
250			13,04	14,846	17,096	18,748
300			13,155	14,826	17,091	18,723
350			13,13	14,857	17,097	18,772
400			13,148	14,863	17,127	18,761
500				14,904	17,089	18,77

Fuente: Hawkins y Zamba (2005).



```
function AED
% Herramienta computacional que permite el análisis de datos y el
% diseño de gráficos de control CUSUM, EWMA y Análisis de Pto de Cambio
% Diseñado por Carolina Pérez y Javier Gómez. Facultad de Ingeniería.
% Universidad de Carabobo. 2012.
%
% Apoyado en el trabajo de "Uso de gráficos EWMA y CUSUM para datos
independientes
% y autocorrelacionados", realizado por las Ingenieras María Méndez y
% Deysi Acosta. Facultad de Ingeniería. Universidad de Carabobo. 2009.
%
% Y
%
% En el trabajo de "Uso del Análisis de Punto de Cambio en conjunto con
los
% gráficos EWMA y CUSUM", realizado por los Ingenieros Domingo
Gutierrez y
% Adriana Ferrer . Facultad de Ingeniería. Universidad de Carabobo.
2010.
%
clc
clear
%
% Lectura de los datos a estudiar
[n, p, yy,Cus]=lec_dat_orig;
cc=max(n);
dd=max(p);
%
% Convirtiendo la matriz en un vector si los datos provienen de una
matriz
if dd>1
    % Creación del vector si p es mayor que 1
    [yy,Cus]=matriz_vector(yy);
end;
hA = runstest(Cus);
if hA==1
    disp(' ')
    disp('ADVERTENCIA: Al aplicar la prueba de aleatoriedad (alfa=5%)')
    disp('se puede decir que hay evidencia suficiente para afirmar')
    disp('que los datos no son aleatorios')
else
    disp(' ')
    disp('Los datos son aleatorios')

end;
datosoriginales=yy;

% Análisis Gráfico antes de la eliminación de outliers
graficos_aed(yy)

%Detección y remoción de outliers de la muestra
```



```
disp(' ')
disp('A continuación se realizará la eliminación de Outliers del vector
original.')
```

```
AUT1=input('¿Autoriza la aplicación de dicha transformación?: Si(1)
No(0) ');
if AUT1==true;
    disp(' ')
    alpha=input('Indique el valor de alpha: ');
    k=input('Indique el valor del número de outliers que espera
encontrar: ');
    [transdat]=eliminar(yy,outliers(yy,alpha,k));
    [yy,Cus]=matriz_vector(transdat);
end;

% Análisis Gráfico después de la eliminación de outliers
graficos_aed_dep(yy)
%
% Estudio de Normalidad
disp(' ')
[h,p] = lillietest(Cus);
if h==1
    disp('Al aplicar la prueba estadística de Lilliefors con un nivel
de')
    disp('confianza del 95% (Alfa=5%), se rechaza la hipótesis que los
datos proceden')
    disp('de una Distribución Normal, dado que p-valor del test
realizado es inferior a 0.05 ')
    disp(' ')
    disp(['El valor de p es: ', num2str(p)])
else
    disp('Al aplicar la prueba estadística de Lilliefors con un nivel
de')
    disp('confianza del 95% (Alfa=5%), no hay suficiente evidencia
estadística')
    disp('para rechazar la hipótesis que los datos proceden de una
Distribución Normal;')
    disp('dado que el p-valor del test realizado es mayor a 0.05')
    disp(' ')
    disp(['El valor de p es: ', num2str(p)])
    disp(' ')
    disp('Advertencia: Se recomienda el uso del gráfico EWMA')
    disp('por ser robusto en la presencia de distribuciones no
Normales,')
    disp('a diferencia del CUSUM y Análisis de punto de cambio.')
```

```
end;
datosdepurados=yy;

% Si hay valores negativos los valores se trasladan de escala para que
todos
% sean positivos y sean admitidos por el programa BOX-COX
% cambio de negativo a positivo si el vector tiene negativos
```



```
cc=length(Cus);
n=cc;
bandn=false;
menv=1000;
for i=1:cc;
    if yy(i)<0 && bandn==false;
        bandn=true;
        menv=min(yy);
    end;
end;
% Formación del vector positivo
if bandn==true;
    gg=zeros([cc 1]);
    for i=1:cc;
        gg(i,1)=yy(i)-menv+0.001;
    end;
    yy=gg;
end
%
%Transformación BOX-COX, para eliminar la no estacionariedad en la
varianza
%Este programa quita la no estacionariedad en la varianza,
%siguiendo la teoría de BOX-COX (1964).
%Consigue el lambda apropiado con respecto a la data la prueba aplicada
%garantiza que la serie resultante es estacionaria en cuanto a la
Varianza.
disp(' ')
disp('A continuación se realizará la transformación de BOX-COX que
elimina')
disp('la no estacionariedad en la varianza.')
AUT1=input('¿Autoriza la aplicación de dicha transformación?: Si(1)
No(0) ');
if AUT1==true;
    [transdat]=boxcox(yy);
    vector_original=transdat;
    Autoriza=input('¿Desea graficar la serie estacionaria en cuanto a
la Varianza?: Si(1) No(0) ');
    if Autoriza==true
        figure('Name','Transformación BOX-COX','NumberTitle','off');
        hold on
        plot(vector_original,'-ro','linewidth',1,...
            'MarkerEdgeColor','k',...
            'MarkerFaceColor','k',...
            'MarkerSize',5)
        xlabel('Tiempo','FontSize',12,'FontName','Andalus')
        ylabel('Datos
Transformados','FontSize',12,'FontName','Andalus')
        title('\fontname{andalus}\fontsize{24}Transformación de BOX-
COX')
        datacursormode on
        hold off
    end;
end;
```



```
end;
if AUT1==true;
    yy=vector_original;
end;
%Análisis Gráfico luego de la transformación Box-Cox
graficos_aed_box(yy)
datosbox=yy;

%
%Evaluación de no estacionariedad en media y transformaciones para
eliminar
%la no estacionariedad en la media, este programa usa los resultados de
las
%pruebas para determinar si los datos son o no estacionarios en cuanto
a la
%media, si no son estacionarios realiza la 1° diferenciación, y si no
son
%estacionarios luego de la 1° diferenciación realizará la 2°
diferenciación
%y nuevamente aplica las pruebas estadísticas. Si aún no son
estacionarias,
%los datos saldrán del sistema.
m=n+1;
vector_original=yy;
%Programa para la aplicación de las pruebas de estacionariedad de
raíces unitarias
%Las Pruebas Estadísticas Aplicadas tienen por nombre Dickey-Fuller y
%Phillips_Perron.
[HDF1]=adftest(vector_original);
[HDF2]=adftest(vector_original);
[HDF3]=adftest(vector_original);
[HPP1]=adftest(vector_original);
[HPP2]=adftest(vector_original);
[HPP3]=adftest(vector_original);
disp(' ')
if(HDF1==1&&HDF2==1&&HDF3==1&&HPP1==1&&HPP2==1&&HPP3==1)
    disp('Los datos son estacionarios en su forma original.')
else
    disp('Los datos no son estacionarios en su forma original.')
    AUT2=input('¿Autoriza la realización de una 1ª diferenciación?:
Si(1) No(0) ');
    if AUT2==true;
        bb=yy;
        bb(2:n+1)=yy;
        bb(1:1)=mean(yy);
        dif1=zeros(1,n);
        for ha=2:m;
            dif1(ha-1)=(bb(ha))-(bb(ha-1));
        end;
        vector_original=dif1;
```





```
apruéb1=input('¿Desea observar los datos diferenciados 1ª vez?:  
Si(1) No(0) ');  
if aprueb1==true  
    figure('Name','Figura','NumberTitle','off');  
    subplot(2,2,1:2);  
    plot(vector_original,'-ro','linewidth',1,...  
         'MarkerEdgeColor','k',...  
         'MarkerFaceColor','k',...  
         'MarkerSize',5)  
    xlabel('Tiempo','FontSize',12,'FontName','Andalus')  
    ylabel('Datos con 1ª  
diferenciación','FontSize',12,'FontName','Andalus')  
    title('\fontname{andalus}\fontsize{18}Datos Transformados -  
1ª diferenciación')  
    datacursormode on  
    hold off  
end;  
end;  
[HDF1]=adftest(vector_original);  
[HDF2]=adftest(vector_original);  
[HDF3]=adftest(vector_original);  
[HPP1]=pptest(vector_original);  
[HPP2]=pptest(vector_original);  
[HPP3]=pptest(vector_original);  
disp(' ')  
if(HDF1==1&&HDF2==1&&HDF3==1&&HPP1==1&&HPP2==1&&HPP3==1)  
disp('Los datos son estacionarios en 1ª diferenciación')  
else  
    disp('Los datos no son estacionarios luego de la 1ª  
diferenciación')  
AUT3=input('¿Autoriza la realización de una 2ª diferenciación?:  
Si(1) No(0) ');  
if AUT3==true;  
    hh=dif1;  
    hh(2:m)=dif1;  
    hh(1:1)=mean(dif1);  
    dif2=zeros(1,n);  
    for ma=3:m;  
        dif2(ma-2)=(hh(ma))-(hh(ma-1))-(hh(ma-1))-(hh(ma-  
2)));  
    end;  
    vector_original=dif2;  
apruéb2=input('¿Desea observar los datos diferenciados 2ª  
vez?: Si(1) No(0) ');  
if aprueb2==true  
    subplot(2,2,3:4);  
    plot(vector_original,'-ro','linewidth',1,...  
         'MarkerEdgeColor','k',...  
         'MarkerFaceColor','k',...  
         'MarkerSize',5)  
    xlabel('Tiempo','FontSize',12,'FontName','Andalus')
```



```
        ylabel('Datos con 2ª
diferenciación', 'FontSize',12, 'FontName', 'Andalus')
        title('\fontname{andalus}\fontsize{18}Datos Transformados -
2ª diferenciación')
        datacursormode on
        hold off
    end;
end;
[HDF1]=adftest(vector_original);
[HDF2]=adftest(vector_original);
[HDF3]=adftest(vector_original);
[HPP1]=pptest(vector_original);
[HPP2]=pptest(vector_original);
[HPP3]=pptest(vector_original);
if (HDF1==1&&HDF2==1&&HDF3==1&&HPP1==1&&HPP2==1&&HPP3==1)
    disp('Los datos son estacionarios en 2º diferenciación')
else
    if AUT3==true;
        disp('Los datos no son estacionarios luego de la 2ª
diferenciación,')
        disp('Los datos no pueden continuar en el sistema.')
```

```
        return
    end;
end;

end;
    %Análisis gráfico luego de las difenciaciones en la media
graficos_aed_dif(vector_original)
end;
hold off
datosdiferenciados=vector_original;
%
% Estudio de Autocorrelación
%
%Aplicación de la prueba H
%Comprobando autocorrelación Prueba h-DURBIN.
%Si el denominador del estadístico h es menor a 0, se aplicara la
prueba
%DURBIN-WATSON original.
disp(' ')
disp('A continuación se aplicará la prueba h-DURBIN para determinar si
los')
disp('datos son o no autocorrelacionados.')
```

```
AUT4=input('¿Autoriza la realización de esta prueba?: Si(1) No(0) ');
disp(' ')
duwa=false;
autoc=0; %se asume autocorrelación
if AUT4==true;
    media=mean(vector_original);
    [nnnnn]=size(vector_original);
    n=max(nnnnn);
    errora=zeros(1,n);
```



```
for i=1:n;
    errora(i)=vector_original(i)-media;
end;
suma1=0;
suma2=0;
for i=2:n;
    suma1=suma1+(errora(i)*errora(i-1));
    suma2=suma2+(errora(i-1)*errora(i-1));
end;
ro=suma1/suma2;
var_est=cov(vector_original);
denominador=n*var_est;
if denominador<1
    duwa=false;
    hdurbin=ro*sqrt(n/(1-n*var_est));
    if((-1.645)<=hdurbin<=(1.645))
        autoc=true;
        disp('Los datos sometidos a evaluación presentan
autocorrelación')
    else
        autoc=false;
        disp('los datos sometidos a evaluación no son
autocorrelacionados')
    end;
else
    disp('No se puede aplicar la prueba h-DURBIN porque el
denominador del')
    disp('estadístico h es menor a 0.')
    disp(' ')
    disp('Se puede realizar la prueba de DURBIN-WATSON sencilla')
    AUT5=input('¿Autoriza la aplicación de dicha prueba?: Si(1)
No(0) ');
    disp(' ')
    if AUT5==true;
        %
        % Prueba de DURBIN-WATSON
        sun=0;
        unos1=vector_original;
        tiemp=vector_original;
        for i=1:n;
            sun=sun+1;
            tiemp(i)=sun;
        end;
        unos1(1:n)=1;
        X11=[unos1, tiemp, vector_original];
        duwa=true;
        [P,DW]=dwtest(errora,X11);
    end;
end;
end;
%
% Determinación del número de datos a estudiar
```



```
dema=vector_original;
[ndat]=size(dema);
ValorN=max(ndat);
%aplicación de prueba DURBIN WATSON
%si el denominador del estadístico h DURBIN es menor que cero
if duwa==true
    %Este programa determina si los datos son o no Autocorrelacionados
    basado
    %en el resultado del estadístico DURBIN-WATSON
    disp('A continuación se leerá el archivo que contiene los valores
de du')
    disp('que son límites de la prueba, para su aplicación.')
    %K representa el número de variables que explican los datos
    %N representa el número de datos a evaluar
    ValorK=4;
    %lectura de data de tabla de durbin watson
    dur=xlsread('DurWat');
    for i=3:34;
        for j=1:1;
            valtab=dur(i,j);
            if valtab==ValorN
                du=dur(i,ValorK+1);
            else
                jax=dur(30:34,1);
                jay=dur(30:34,2);
                du=spline(jax,jay,ValorN);
            end;
        end;
    end;
    end;
    disp(' ')
    if (0<DW)&&(DW<du)
        autoc=true;
        disp('Los datos presentan autocorrelación positiva')
    else
        if ((4-du<DW)&&(DW<4))
            autoc=true;
            disp('Los datos presentan autocorrelación negativa')
        else
            if ((du<DW)&&(DW<4-du))
                autoc=false;
                disp('Los datos no son autocorrelacionados')
            end;
        end;
    end;
end;

end;

if autoc==false
    disp(' ')
end;
```



```
disp('A continuación se diseñarán los Gráficos de Control.')
```

```
disp('Para el Diseño del Gráfico EWMA con el Análisis de Punto')
```

```
disp('de Cambio, seleccione (1); y para el Diseño del Gráfico CUSUM
```

```
con')
```

```
AUT6=input('el Análisis de Punto de Cambio, seleccione (2): ');
```

```
disp(' ')
```

```
if AUT6==true;
```

```
%
```

```
% Construcción de gráfico EWMA
```

```
disp(['A continuación se mostrará la media de los datos: ',
```

```
num2str(mean(Cus))])
```

```
disp('D Significa la magnitud del cambio en la media que debe ser
```

```
detectada')
```

```
disp('lo más rápido posible.')
```

```
disp(' ')
```

```
disp('Valores de D permitidos')
```

```
disp('0 0.25 0.5 0.75 1 1.25 1.5 1.75 2 2.25 2.5 2.75 3 3.25 3.5 3.75
```

```
4')
```

```
dcambio=input('Indique el valor de D: ');
```

```
%ARLewma=370
```

```
%lectura de la tabla ewma
```

```
%lectura de data de tabla de EWMA
```

```
disp(' ')
```

```
disp('A continuación se leerá la tabla que contiene los valores de
```

```
Lamda y K.')
```

```
ValEwmaI = xlsread('EWMA_1');
```

```
[newi,pewi]=size(ValEwmaI);
```

```
%obtención de los valores de lamda y K
```

```
for i=3:19;
```

```
    valDew=ValEwmaI(i,1);% j=1
```

```
        if valDew==dcambio
```

```
            j1=2;
```

```
            encv=false;
```

```
            mpewi=max(pewi);
```

```
            while ((encv==false)&&(j1<=mpewi))
```

```
                j1=(j1)+1;
```

```
                anter=ValEwmaI(i,((j1)-1));
```

```
                presen=ValEwmaI(i,j1);
```

```
                if anter<=presen % falta un =?
```

```
                    land=ValEwmaI(1,j1-1);
```

```
                    kcons=ValEwmaI(i,1);
```

```
                    encv=true;
```

```
                end;
```

```
            end;
```

```
        end;
```

```
    end;
```

```
end;
```

```
%cálculo de los puntos a graficar en EWMA
```

```
nn=max(ndat);
```

```
zetal(1:1)=mean(Cus);
```

```
for i10=1:nn;
```

```
    zetal(i10+1)=(1-land)*zetal(i10)+(land)*Cus(i10);
```

```
end;
```



```
mediaei=mean(Cus);
desvei=std(Cus);
%límites de control del gráfico EWMA
LCsewma=mediaei+kcons*(desvei)*sqrt(land/(2-land));
LCiewma=mediaei-kcons*(desvei)*sqrt(land/(2-land));
LCewma=mediaei;
%vectores a graficar
LimSupE(1:nn)=LCsewma;
LimInfE(1:nn)=LCiewma;
LimCentE(1:nn)=LCewma;
disp(' ')
grafewm=input('¿Desea ver el gráfico EWMA? Si(1) No(0) ');
if grafewm==true
    figure('Name','Gráfico de Control EWMA','NumberTitle','off');
    hold on
    plot(LimSupE,'-b','linewidth',2,...
        'MarkerEdgeColor','b',...
        'MarkerFaceColor','b',...
        'MarkerSize',5)
    plot(LimCentE,'-r','linewidth',1,...
        'MarkerEdgeColor','r',...
        'MarkerFaceColor','r',...
        'MarkerSize',5)
    plot(LimInfE,'-b','linewidth',2,...
        'MarkerEdgeColor','b',...
        'MarkerFaceColor','b',...
        'MarkerSize',5)
    plot(zeta1,'-mo','linewidth',1,...
        'MarkerFaceColor',[0.24705882370472 0.24705882370472
0.24705882370472],...
        'MarkerEdgeColor',[0.24705882370472 0.24705882370472
0.24705882370472],...
        'MarkerSize',5)

    xlabel('Tiempo','FontSize',14,'FontName','Andalus')
    ylabel('EWMA','FontSize',14,'FontName','Andalus')
    title('\fontname{andalus}\fontsize{18}Gráfico de Control
EWMA')
    legend('Límite Superior','Límite Central','Límite
Inferior','Z',0)
    datacursormode on
    hold off
end;
else
    %
    % Construcción del gráfico CUSUM
    hcusum=4;
    medcusum=mean(Cus);
    desvcus=std(Cus);
    disp(' ')
    disp('A continuación debe elegir el número de desviaciones típicas que
determinará')
```



```
disp('la sensibilidad del gráfico (mientras más pequeño sea este
número, más ')
disp('sensible es el gráfico a pequeños cambios).')
disp('El valor debe ser mayor que 0 y puede usar decimales')
Ndesv=input('Indique el N° de desviaciones: ');
medcus1=medcusum-(Ndesv*desvcus);
kcusum=(medcusum-medcus1)/2;
banc1=false;
banc2=false;
%límites de control del gráfico CUSUM
LCScusum=medcusum+(hcsusum*desvcus);
LCIcusum=medcusum-(hcsusum*desvcus);
LCcusum=medcusum;
%Proceso de elaboración de los vectores de CUSUM
nn=max(ValorN);
i=1;
fuecontpos=true;
fuecontneg=true;
Cpos(1:1)=medcusum;
Cneg(1:1)=medcusum;
Npos=0;
Nneg=0;
vecNpos=zeros(1,nn);
vecNneg=zeros(1,nn);
while ((fuecontpos==true)&&(fuecontneg==true)&&(i<nn));
    i=i+1;
    CUMpos=Cpos(i-1)+(Cus(i-1)-medcusum-kcusum);
    CUMneg=Cneg(i-1)+(Cus(i-1)-medcusum+kcusum);
    %desviación positiva
    if (CUMpos)>=LCcusum;
        Cpos(i)=CUMpos;
    else
        Cpos(i)=LCcusum;
        Npos=0;
        vecNpos(i)=Npos;
    end;
    if CUMpos>LCScusum;
        Npos=Npos+1;
        vecNpos(i)=Npos;
        if (Cpos(i)>=LCScusum);

            Fuera_de_control=i;
            fuecontpos=0;
            banc1=true;
        end;
    end;
    %desviaciones negativas
    if (CUMneg)<LCcusum;
        Cneg(i)=CUMneg;
    else
        Cneg(i)=LCcusum;
        Nneg=0;
```



```
        vecNneg(i)=Nneg;
    end;
    if CUMneg<LCcusum;
        Nneg=Nneg+1;
        vecNneg(i)=Nneg;
        if (Cneg(i)<=LCIcusum);

            Fuera_de_control=i;
            fuecontneg=0;
            banc2=true;
        end;
    end;
end;
%cambio indica la posición del vector en la que ocurrió el cambio
if banc1==true
    cambio=Fuera_de_control-(Npos-1);
else
    if banc2==true
        cambio=Fuera_de_control-(Nneg-1);
    else
        disp(' ')
        disp('El proceso está bajo control')
    end;
end;
%Terminación del vector en caso de que salga de control esta operación
se
%realiza para efectos de la gráfica.
if (banc1==true)|| (banc2==true)
    for i=Fuera_de_control:nn;
        CUMpos=Cpos(i-1)+(Cus(i-1)-medcusum)-kcusum;
        CUMneg=Cneg(i-1)+(Cus(i-1)-medcusum)+kcusum;
        if (CUMpos)>=LCcusum;
            Cpos(i)=CUMpos;
        else
            Cpos(i)=LCcusum;
        end;
        if CUMneg<LCcusum;
            Cneg(i)=CUMneg;
        else
            Cneg(i)=LCcusum;
        end;
    end;
end;
%vectores a graficar del CUSUM
LimSupC(1:nn)=LCScusum;
LimInfC(1:nn)=LCIcusum;
LimCentC(1:nn)=LCcusum;
if (banc1==true)|| (banc2==true)
    disp(' ')
    disp('A continuación se mostrará la posición a partir de la cual el
proceso')
    disp(['se salió de control: ', num2str(Fuera_de_control)])
```





```
disp(' ')
disp('A continuación se mostrará la posición a partir de la cual
ocurrió el')
disp(['cambio que hizo que los datos se salieran de control: ',
num2str(cambio)])
end;
disp(' ')
grafcus=input('¿Desea ver el gráfico de control CUSUM?: Si(1) No(0) ');
if grafcus==true
figure('Name','Gráfico de Control CUSUM','NumberTitle','off');
hold on
plot(LimSupC,'-b','linewidth',2,...
'MarkerEdgeColor','b',...
'MarkerFaceColor','b',...
'MarkerSize',5)
plot(LimCentC,'-r','linewidth',1,...
'MarkerEdgeColor','r',...
'MarkerFaceColor','r',...
'MarkerSize',5)
plot(LimInfC,'-b','linewidth',2,...
'MarkerEdgeColor','b',...
'MarkerFaceColor','b',...
'MarkerSize',5)
plot(Cpos,'-ko','linewidth',1,...
'MarkerEdgeColor','k',...
'MarkerFaceColor','k',...
'MarkerSize',5)
plot(Cneg,'-mo','linewidth',1,...
'MarkerEdgeColor','m',...
'MarkerFaceColor','m',...
'MarkerSize',5)
xlabel('Tiempo','FontSize',14,'FontName','Andalus')
ylabel('CUSUM','FontSize',14,'FontName','Andalus')
title('\fontname{andalus}\fontsize{18}Gráfico de Control
CUSUM')
legend('Límite Superior','Límite Central','Límite
Inferior','Cpos','Cneg',0)
datacursormode on
hold off
end;
end;
%
%Aplicación de Análisis de Punto de Cambio
%Se empieza a controlar luego de 9 mediciones (datos)
%Determinación de Gmaxn
%
Gmaxn=GMAXN( vector_original,cc );
%Lectura del estadístico de comparación h
disp(' ')
disp('A continuación se procederá a construir el Gráfico de Control
de')
disp('Análisis de Punto de Cambio.')
```



```
disp(' ')
disp('Deberá indicar el valor del alfa (probabilidad especificada de
emitir una señal errónea)')
disp('Valores permitidos de alfa: ')
disp('0.001 0.002 0.005 0.01 0.02 0.05')
alfa=input('Indique el valor de alfa: ');
limH=TabH(cc,alfa);
%comparación del estadístico
Donde=false;
for P1=10:n
    if Gmaxn(P1)>=limH(P1)&& Donde==false
        Pk=punto_de_cambio(vector_original,P1);
        disp(' ')
        disp(['Se ha detectado un cambio en la media o varianza en: ',
num2str( P1 )]) %Donde se detecta el cambio.
        disp(['El punto donde ocurre el cambio es: ', num2str(Pk)])
%Donde ocurre el cambio detectado.
        %¿cuál es el que cambia media o varianza?
        %Para probar el cambio en la media se utilizará la prueba t
para 2
        %muestras que proporciona MatLab
aux01=ttest2(vector_original(1:Pk),vector_original(Pk+1:P1));
        %Para probar el cambio en la varianza se utilizará la prueba F
para 2
        %muestras que proporciona MatLab
aux02=varTest2(vector_original(1:Pk),vector_original(Pk+1:P1));
        if aux01==true
            if aux02==true
                disp(' ')
                disp('Existe evidencia estadística de un cambio en la
media y en la varianza')
                disp(['con un ',num2str( 100-(alfa*100) ),'% de
confianza'])
                disp(['La Media para los datos antes del cambio es: ',
num2str(Xik(Cus,0,Pk))])
                disp(['La Media para los datos después del cambio es:
', num2str(Xik(Cus,Pk,cc))])
                disp(['La Varianza para los datos antes del cambio es:
',num2str((Vik(Cus,0,Pk))/(Pk-1))])
                disp(['La Varianza para los datos después del cambio
es: ',num2str((Vik(Cus,Pk,cc))/(cc-Pk-1))])
            else
                disp(' ')
                disp('Existe evidencia estadística de un cambio en la
media')
                disp(['con un ',num2str( 100-(alfa*100) ),'% de
confianza'])
                disp(['La Media para los datos antes del cambio es: ',
num2str(Xik(Cus,0,Pk))])
                disp(['La Media para los datos después del cambio es:
', num2str((Xik(Cus,Pk,cc)))]])
            end
        end
    end
end
```



```
disp(['La Varianza para los datos es: ',num2str(((Vik(Cus,0,Pk))+(Vik(Cus,Pk,cc)))/(n-2))])
end
else
    if aux02==true
        disp(' ')
        disp('Existe evidencia estadística de un cambio en la
varianza')
        disp(['con un ',num2str(100-(alfa*100)), '% de
confianza'])
        disp(['La Media para los datos es: ',
num2str(Xik(Cus,0,cc))])
        disp(['La Varianza para los datos antes del cambio es: ',
num2str((Vik(Cus,0,Pk))/(Pk-1))])
        disp(['La Varianza para los datos después del cambio
es: ',num2str((Vik(Cus,Pk,cc))/(cc-Pk-1))])
    end
    end
    Donde=true;
end;
end;
if Donde==false
    disp(' ')
    disp('No existe evidencia estadística para afirmar que ocurrió un
cambio')
    disp('en la media o la varianza')
    disp(' ')
    disp('El proceso se encuentra bajo control')
    disp(' ')
    disp(['La Media para los datos es: ', num2str(Xik(Cus,0,cc))])
    disp(['La Varianza para los datos es: ',
num2str((Vik(Cus,0,cc))/(cc-1))])
end;
%gráfica del Análisis de Punto de Cambio
figure('NumberTitle','off','Name','Análisis de Punto de Cambio');
hold on
plot(limH,'-b','linewidth',2,...
'MarkerEdgeColor','b',...
'MarkerFaceColor','b',...
'MarkerSize',5)
plot(Gmaxn,'-ko','linewidth',1,...
'MarkerEdgeColor','r',...
'MarkerFaceColor','r',...
'MarkerSize',5)
xlim([10 n])
xlabel('Muestras','FontSize',14,'FontName','Andalus')
ylabel('Gmax,n','FontSize',14,'FontName','Andalus')
title('\fontname{andalus}\fontsize{18}Gráfico de Control de
Punto de Cambio')
legend('Límite de Control h','Gmax,n',0)
datacursormode on
hold off
```



```
else

    %Gráficos de control para datos independientes
    %en el caso de decidir obviar el fenómeno de autocorrelación
    %para ilustrar las variaciones si esto ocurriese

    disp(' ')
    ignauto = input('¿Desea ignorar el fenómeno de autocorrelacion?:
Si(1) No(0) ')
    if ignauto==true

        %Determinación del modelo de serie de tiempo

        Coef_Auto
        if fffg==true;
            return
        end;

        %Determinación de los parámetros del modelo

        para_metros;

        %Consiguiendo algunos parámetros a partir de los anteriores
        %mediante una fórmulas

        formulas;

        %Gráficos de control para datos AUTOCORRELACIONADOS

        EWMA_auto;
        CUSUM_autoo;
        end
    end

end

end
```



```
function APC_prog
%programa para la utilización de Análisis de Punto de Cambio, como
grafico
%de control de la calidad
%Se empieza a controlar luego de 9 mediciones (datos)
%Determinación de Gmaxn
for Z1=10:n;
    Gmax=0;
    for k=1:Z1;
        Gkn=((k*log((Vik(yy,0,Z1)/(Z1))/(Vik(yy,0,k)/(k)))+(Z1-
k)*log((Vik(yy,0,Z1)/(Z1))/(Vik(yy,0,k)/(k)))/(1+(11/12)*(1/k+1/(Z1-
k)-1/Z1)+(1/(k^2)+1/((Z1-k)^2)-1/(Z1^2)))));
        if Gkn>Gmax;
            Gmax=Gkn;
        end;
    end;
    Gmaxn(Z1)=Gmax;
end;
Gmaxn
%lectura del estadístico de comparación h
disp('A continuación debe indicar el alfa (el error de emitir una señal
errónea)')
disp('Valores permitidos de alfa: ')
disp('0.001 0.002 0.005 0.01 0.02 0.05')
alfa=input('Indique el valor de alfa: ','s');
TabAPC=xlsread('TabHawkins');
[nAPC,pAPC]=size(TabAPC);
H=zeros(n,1);
for N1=10:n;
    if N1>14
        if (0.001<=alfa<0.05)
            h=1.58-2.52*log(alfa)+(0.094+.33*log(alfa)/(sqrt(N1-9)));
            H(N1)=h;
        end;
        if (alfa==0.05)
            h=8.43+.074*log(N1-9);
            H(N1)=h;
        end;
    else
        for i=3:nAPC;
            ValAPC=TabAPC(i,1);
            if ValAPC==N1;
                for j=2:pAPC;
                    Val2APC=TabAPC(2,j);
                    if Val2APC==alfa;
                        h=TabAPC(i,j);
                        H(N1)=h;
                    end;
                end;
            end;
        end;
    end;
end;
```



```
end;
%comparación del estadístico
Donde=false;
for P1=10:n
    if Gmaxn(P1)>=H(P1)&& Donde==false
        Pos=P1;
        disp('Existe evidencia de un cambio en la media o varianza.',
num2str(Pos) )
        %cual es el que cambia media o varianza
        Donde=true;
    end;
end;
%gráfica del APC
hold on
    plot(H, '-b', 'linewidth',1,...
        'MarkerEdgeColor','b',...
        'MarkerFaceColor','b',...
        'MarkerSize',5)
    plot(Gmaxn, '-ro', 'linewidth',1,...
        'MarkerEdgeColor','r',...
        'MarkerFaceColor','r',...
        'MarkerSize',5)
    xlabel('Tiempo')
    ylabel('Análisis de Punto de Cambio')
    title('\fontname{andalus}Gráfico de Control
CUSUM', 'FontSize',14)
    legend('límite superior', 'límite central', 'límite
inferior', '',0)
    hold off
end
```

```
function [ c ] = C_Bartlett( k,n )
%programa para determinar la constante de Bartlett
c=(1+(11/12)*(((1)/(k))+((1)/(n-k))-((1)/(n))))+(1/(k^2))+1/((n-
k)^2))-1/(n^2));
end
```

```
%Determinación del modelo de serie de tiempo
```

```
%Determinación de coeficientes de autocorrelación
%y determinación de coeficientes de autocorrelación parcial
```

```
[coefauto lags bounds ] = autocorr(dema);
```



```
[coefpar lags bounds ] = parcorr(dema);
[fa]=size(coefauto);
[ga]=size(coefpar);
ESAR=false;
ESARMA=false;
fffg=false;
disp('A continuación se mostrará el gráfico de los coeficientes de
autocorrelación')
aprueb3 = input('Autoriza mostrar dicha gráfica? Si(1) No(0): ')
if aprueb3==true
    autocorr(dema)
end;
AutoAR = input(';¿Los coeficientes descienden rápidamente de forma
exponencial y/o sinusoidal? Si(1) No(0): ')
disp('A continuación se mostrará el gráfico de los coeficientes de
autocorrelación parcial')
vicoef = input('Autoriza mostrar dicha gráfica? Si(1) No(0): ')
if vicoef==true
    parcorr(dema)
end;
AutoMA = input(';¿Los coeficientes descienden rápidamente de forma
exponencial y/o sinusoidal? Si(1) No(0): ')
if (AutoAR==true)&(AutoMA==false)
    modeloAR=true;
    modeloMA=false;
    modeloARMA=false;
    disp('La serie sigue un modelo AR')
    ESAR=true;
end;
if (AutoAR==false)&(AutoMA==true)
    modeloMA=true;
    modeloAR=false;
    modeloARMA=false;
    disp('La serie sigue un modelo MA ')
    disp('Los datos no pueden continuar en el sistema')
    return;
end;
if (AutoAR==true)&(AutoMA==true)
    modeloARMA=true;
    modeloAR=false;
    modeloMA=false;
    disp('La serie sigue un modelo ARMA')
    ESARMA=true;
end;
if (AutoAR==false)&(AutoMA==false)
    modeloARMA=false;
    modeloAR=false;
    modeloMA=false;
    disp('La serie sigue un modelo que no puede ser identificado')
    disp('por este programa ya que no sigue un modelo ARIMA')
    disp('Los datos deben salir del sistema')
    fffg=true;
```



```
    return;
end;

%Gráfico CUSUM para datos autocorrelacionados

% Instrucciones para seleccionar valores de Phi Y Shi tabulados

disp('Dado el valor de Phil de las fórmulas se debe elegir')
disp('el valor tabulado más cercano, el valor real es: ')
Phil
disp('Puede elegir Phi = 0.4 ó 0.8')
Phi=input('Diga el valor de Phi: ')
disp('Dado el valor de Shi de las fórmulas se debe elegir')
disp('El valor tabulado más cercano, el valor real es: ')
Shil
disp('Puede elegir Shi = 0.5 ó 0.9')
Shi=input('Diga el valor de Shi: ')

%Instrucciones para elegir el valor de cambio de la media tabulado

desvaut=desvx;
disp('A continuación vera la media de los datos transformados')

MEDIAA=mean(Cus)

disp('Debe elegir entre los siguientes valores que estan')
disp('tabulados el valor del cambio en la media que usted quiere')
disp('detectar más rápidamente')
disp('0,5    1    1,5    2    2,5    3    4    5')
cambt=input('Indique el valor seleccionado: ')

%Obtención de r de la tabla
i=0;
j=0;
Lec_Dat_CUA;
menr=false;
nmo=3;
for i=1:1;
    vall1=ValCusumA(i,3);
    if vall1==Phi;
        for j=2:2;
            vall2=ValCusumA(j,3);
```





```
if val12==Shi;
    k=0;
    l=0;
    ll=0;
    for k=3:3;
        for l=3:10;
            val13=ValCusumA(k,l);
            if val13==cambt;
                while (menr==false)&(nmo<17);
                    nmo=nmo+1;
                    val14=ValCusumA(nmo,l);
                    val15=ValCusumA(nmo+1,l);
                    if val14<=val15;
                        erre=ValCusumA(nmo,2);
                        menr=true;
                    end;
                end;
            end;
        end;
    end;
else
    val12=ValCusumA(19,3);
    if val12==Shi;
        k=0;
        l=0;
        ll=0;
        menr=false;
        nmo=20;
        for k=20:20;
            for l=3:10;
                val13=ValCusumA(k,l);
                if val13==cambt;
                    while (menr==false)&(nmo<34);
                        nmo=nmo+1;
                        val14=ValCusumA(nmo,l);
                        val15=ValCusumA(nmo+1,l);
                        if val14<=val15;
                            erre=ValCusumA(nmo,2);
                            menr=true;
                        end;
                    end;
                end;
            end;
        end;
    end;
else
    val11=ValCusumA(35,3);
    if val11==Phi;
        menr=false;
        for j=36:36;
```



```
val12=ValCusumA(j,3);
if val12==Shi;
    k=0;
    l=0;
    ll=0;
    nmo=37;
    for k=37:37;
        for l=3:10;
            val13=ValCusumA(k,l);
            if val13==cambt;
                while (menr==false)&(nmo<51);
                    nmo=nmo+1;
                    val14=ValCusumA(nmo,l);
                    val15=ValCusumA(nmo+1,l);
                    if val14<=val15;
                        erre=ValCusumA(nmo,2);
                        menr=true;
                    end;
                end;
            end;
        end;
    end;
else
    val12=ValCusumA(53,3);
    if val12==Shi;
        k=0;
        l=0;
        ll=0;
        menr=false;
        nmo=54;
        for k=54:54;
            for l=3:10;
                val13=ValCusumA(k,l);
                if val13==cambt;
                    while (menr==false)&&(nmo<68);
                        nmo=nmo+1;
                        val14=ValCusumA(nmo,l);
                        val15=ValCusumA(nmo+1,l);
                        if val14<=val15;
                            erre=ValCusumA(nmo,2);
                            menr=true;
                        end;
                    end;
                end;
            end;
        end;
    end;
end;
end;
end;
end;
end;
end;
end;
```



```
erre

disp('Debe elegir entre los valores que estan')
disp('tabulados el valor r que sea más cercano al erre obtenido')
disp('para efectos de buscar en la tabla de los C')
disp('0.25    0.5    1')
errel=input('Indique el valor seleccionado: ')

% Instrucciones para seleccionar valores de phi y shi tabulados

disp('Dado el valor de Phil de las fórmulas se debe elegir')
disp('el valor tabulado más cercano, el valor real es: ')
Phil
disp('Puede elegir Phi = 0.1    0.2    0.3    0.4    0.5    0.6    0.7    0.8
0.9')
Phi=input('Diga el valor de Phi: ')
disp('Dado el valor de Shil de las fórmulas se debe elegir')
disp('El valor tabulado más cercano, el valor real es: ')
Shil
disp('Puede elegir Shi = 0.1    0.2    0.3    0.4    0.5    0.6    0.7    0.8
0.9')
Shi=input('Diga el valor de Shi: ')

%Obtención del valor de c de la tabla dependiendo de Phi, Shi Y r

Lec_Dat_CUS_C

vall1=VECC(1,2);
if vall1==errel;
    for i=2:10;
        valm=VECC(2,i);
        if valm==Shi;
            for j=3:11;
                valn=VECC(j,1);
                if valn==Phi;
                    ceec=VECC(j,i);
                end;
            end;
        end;
    end;
else
    valo=VECC(12,2);
    if valo==errel;
        for i=2:10;
            valm=VECC(13,i);
            if valm==Shi;
                for j=14:22;
                    valn=VECC(j,1);
                    if valn==Phi;
```



```

        ceee=VECC(j,i);
    end;
end;
end;
else
    valp=VECC(23,2);
    if valp==errel;
        for i=2:10;
            valm=VECC(24,i);
            if valm==Shi;
                for j=25:33;
                    valn=VECC(j,1);
                    if valn==Phi;
                        ceee=VECC(j,i);
                    end;
                end;
            end;
        end;
    end;
end;
end;
end;
ceee;

%Límites de control del gráfico CUSUM para datos autocorrelacionados

LCScusuma=mean(Cus)+ceee*desvaut;
LCIcusuma=mean(Cus)-ceee*desvaut;
LCcusuma=mean(Cus);

%Proceso de elaboración de los vectores CUSUM de datos
autocorrelacionados
medcusuma=mean(Cus);
i=1;
fuecontposa=true;
fuecontnega=true;
Cposa(1:1)=medcusuma;
Cnega(1:1)=medcusuma;
Nposa=0;
Nnega=0;
banc11=false;
banc22=false;
nn=max(ValorN);
while ((fuecontposa==true)&(fuecontnega==true))&(i<=nn);
    i=i+1;
    CUMposa=Cposa(i-1)+(Cus(i-1)-medcusuma-erre*desvaut);
    CUMnega=Cnega(i-1)+(Cus(i-1)-medcusuma+erre*desvaut);

%Desviaciones positivas
```



```
if (CUMposa)>=LCcusuma;
    Cposa(i)=CUMposa;
else
    Cposa(i)=LCcusuma;
    Nposa=0;
    vecNposa(i)=Nposa;
end;
if CUMposa>LCcusuma;
    Nposa=Nposa+1;
    vecNposa(i)=Nposa;
    if (Cposa(i)>=LCScusuma);
        POSposa= Nposa;
        Fuera_de_Controla=i;
        fuecontposa=false;
        banc11=true;
    end;
end;

%Desviaciones negativas
if CUMnega<LCcusuma;
    Cnega(i)=CUMnega;
else
    Cnega(i)=LCcusuma;
    Nnega=0;
    vecNnega(i)=Nnega;
end;
if CUMnega<LCcusuma;
    Nnega=Nnega+1;
    vecNnega(i)=Nnega;
    if (Cnega(i)<=LCIcusuma);
        POSnega= Nnega;
        Fuera_de_Controla=i;
        fuecontnega=0;
        banc22=true;
    end;
end;

%cambio indica la posición del vector en la que ocurrió el cambio
if banc11==true
    cambioo=Fuera_de_Controla-(Nposa-1);
end;
if banc22==true
    cambioo=Fuera_de_Controla-(Nnega-1);
end;
if(banc11==false)&(banc22==false)
    disp('El proceso está bajo control')
end;
```



```
%Terminación del vector en el caso de que se salga de control
%esta operación se realiza para efectos de la gráfica

if (banc11==true)|(banc22==true)
    for i=Fuera_de_Controla:nn+1;
        CUMAposa=Cposa(i-1)+(Cus(i-1)-medcusuma-erre*desvaut);
        CUMAnega=Cnega(i-1)+(Cus(i-1)-medcusuma+erre*desvaut);
        if (CUMposa)>LCcusuma;
            Cposa(i)=CUMposa;
        else
            Cposa(i)=LCcusuma;
        end;
        if CUMnega<LCcusuma;
            Cnega(i)=CUMnega;
        else
            Cnega(i)=LCcusuma;
        end;
    end;
end;

%Vectores a graficar del CUSUM independiente

LimSupCa(1:nn)=LCScusuma;
LimInfCa(1:nn)=LCIcusuma;
LimCentCa(1:nn)=LCcusuma;
Cposa;
Cnega;
if (banc11==true)|(banc22==true)
    disp('El siguiente valor indica la posición en la que los')
    disp('datos comenzaron a estar')
    Fuera_de_Controla
    disp('El siguiente valor indica la posición en la que')
    disp('ocurrió un ')
    cambioo
end;
grafcusa=input('Diga true para mostrar el gráfico CUSUM para datos
autocorrelacionados: ')
if grafcusa==true
    hold on

    plot(LimSupCa, '-b', 'LineWidth',1,...
        'MarkerEdgeColor', 'r',...
        'MarkerFaceColor', 'r',...
        'MarkerSize',5)

    plot(LimCentCa, '-K', 'LineWidth',1,...
        'MarkerEdgeColor', 'r',...
```



```
'MarkerFaceColor','r',...
'MarkerSize',5)

plot(LimInfCa, '-b', 'LineWidth',1,...
'MarkerEdgeColor','r',...
'MarkerFaceColor','r',...
'MarkerSize',5)

plot(Cposa, '-ro', 'LineWidth',1,...
'MarkerEdgeColor','r',...
'MarkerFaceColor','r',...
'MarkerSize',5)

plot(Cnega, '-mo', 'LineWidth',1,...
'MarkerEdgeColor','m',...
'MarkerFaceColor','m',...
'MarkerSize',5)

xlabel('Tiempo')
ylabel('Cusum ')
title('\fontname{andalus}CUSUM PARA DATOS
AUTOCORRELACIONADOS', 'FontSize',14)

legend('Límite Superior', 'Límite Central', 'Límite
Inferior', 'Cposa', 'Cnega',0)
hold off
end;

function newy = eliminar(y,t)
y = y(:);
%% calculate statistics for up to k outliers
if t==0
    newy=y;
else
for i = 1:t
ys = abs(y-mean(y))/std(y);
posmax=1;
for a=2:length(y)
    if ys(a)>ys(posmax)
        posmax=a;
    end
end
r(i)=ys(posmax,1);
yyy=[];
m=0;
for l= 1:length(y)
```



```
        if l~=posmax
            m=m+1;
            yyy(m,1)=y(l);
        end
    end;
end;
y = [];
for l = 1 :length(yyy)
y(l,1) = yyy(l,1);
end;
newy=y;
end
end

%Gráfico EWMA para datos autocorrelacionados

% Instrucciones para seleccionar valores de Phi Y Shi tabulados

disp('Dado el valor de Phil de las fórmulas se debe elegir')
disp('el valor tabulado más cercano, el valor real es: ')
Phil
disp('Puede elegir Phi = 0.4 ó 0.8')
Phi=input('Diga el valor de Phi: ')
disp('Dado el valor de Shil de las formulas se debe elegir')
disp('El valor tabulado más cercano,el valor real es: ')
Shil
disp('Puede elegir Shi = 0.5 ó 0.9')
Shi=input('Diga el valor de Shi: ')

%Instrucciones para elegir el valor de cambio de la media tabulado

desvaut=desvx;
disp('A continuación vera la media de los datos transformados')

MEDIAA=mean(Cus)

disp('Debe elegir entre los siguientes valores que estan')
disp('tabulados el valor del cambio en la media que usted quiere')
disp('detectar más rapidamente')
disp('0,5    1    1,5    2    2,5    3    4    5')
cambt=input('Indique el valor seleccionado: ')

%Obtención de LAMDA de la tabla
i=0;
j=0;
```





```
Lec_Dat_EWA;
menr=false;
nmo=3;
for i=1:1;
    val11=ValEwmaA(i,4);
    if val11==Shi;
        for j=2:2;
            val12=ValEwmaA(4,j);
            if val12==Phi;
                k=0;
                l=0;
                ll=0;
                for k=2:2;
                    for l=4:11;
                        val13=ValEwmaA(k,l);
                        if val13==cambt;
                            while (menr==false)&(nmo<13);
                                nmo=nmo+1;
                                val14=ValEwmaA(nmo,l);
                                val15=ValEwmaA(nmo+1,l);
                                if val14<=val15;
                                    lamdat=ValEwmaA(nmo,3);
                                    menr=true;
                                end;
                            end;
                        end;
                    end;
                end;
            end;
        else
            k=0;
            l=0;
            ll=0;
            menr1=false;
            nmo=18;
            for k=17:17;
                for l=4:11;
                    val13=ValEwmaA(k,l);
                    if val13==cambt;
                        while (menr1==false)&(nmo<28);
                            nmo=nmo+1;
                            val14=ValEwmaA(nmo,l);
                            val15=ValEwmaA(nmo+1,l);
                            if val14<=val15;
                                lamdat=ValEwmaA(nmo,3);
                                menr1=true;
                            end;
                        end;
                    end;
                end;
            end;
        end;
    end;
end;
```



```
else
  for j=2:2;
    val12=ValEwmaA(34,j);
    if val12==Phi;
      k=0;
      l=0;
      ll=0;
      menr2=false;
      nmo=33;
      for k=32:32;
        for l=4:11;
          val13=ValEwmaA(k,l);
          if val13==cambt;
            while (menr2==false)&(nmo<43);
              nmo=nmo+1;
              val14=ValEwmaA(nmo,l);
              val15=ValEwmaA(nmo+1,l);
              if val14<=val15;
                lamdat=ValEwmaA(nmo,3);
                menr2=true;
              end;
            end;
          end;
        end;
      end;
    else
      k=0;
      l=0;
      ll=0;
      menr3=false;
      nmo=49;
      for k=48:48;
        for l=4:11;
          val13=ValEwmaA(k,l);
          if val13==cambt;
            while (menr3==false)&(nmo<59);
              nmo=nmo+1;
              val14=ValEwmaA(nmo,l);
              val15=ValEwmaA(nmo+1,l);
              if val14<=val15;
                lamdat=ValEwmaA(nmo,3);
                menr3=true;
              end;
            end;
          end;
        end;
      end;
    end;
  end;
end;
lamdat
```



```
%Obtención del valor de C de la tabla dependiendo de Phi, Shi Y LAMDA

disp('Debe elegir entre los siguientes valores que estan')
disp('tabulados el valor "lamda" que sea más cercano al LAMDA
obtenido')
disp('para efectos de buscar en la tabla de los C')
disp('0.1    0.2    0.3')
LAMDA=input('Indique el valor seleccionado: ')

% Instrucciones para seleccionar valores de Phi Y Shi tabulados

disp('Dado el valor de Phil de las fórmulas se debe elegir')
disp('el valor tabulado más cercano , el valor real es: ')
Phil
disp('Puede elegir Phi = 0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8
0.9')
Phi=input('Diga el valor de Phi: ')
disp('Dado el valor de Shi de las fórmulas se debe elegir')
disp('El valor tabulado más cercano, el valor real es: ')
Shil
disp('Puede elegir Shi = 0.1  0.2  0.3  0.4  0.5  0.6  0.7  0.8
0.9')
Shi=input('Diga el valor de Shi: ')

Lec_Dat_EWA_C

valll=VEC(1,2);
if valll==LAMDA;
    for i=2:10;
        valm=VEC(2,i);
        if valm==Shi;
            for j=3:11;
                valn=VEC(j,1);
                if valn==Phi;
                    ceat=VEC(j,i);
                end;
            end;
        end;
    end;
else
    valo=VEC(12,2);
    if valo==LAMDA;
        for i=2:10;
            valm=VEC(13,i);
            if valm==Shi;
                for j=14:22;
                    valn=VEC(j,1);
```



```
                if valn==Phi;
                    ceat=VEC(j,i);
                end;
            end;
        end;
    end;
else
    valp=VEC(23,2);
    if valp==LAMDA;
        for i=2:10;
            valm=VEC(24,i);
            if valm==Shi;
                for j=25:33;
                    valn=VEC(j,1);
                    if valn==Phi;
                        ceat=VEC(j,i);
                    end;
                end;
            end;
        end;
    end;
end;
end;
ceat;

%Obtención del vector EWMA

rr=0;
zewaut(1:1)=mean(Cus);
mon=ValorN+1;
for rr=1:ValorN;
    zewaut(rr+1)=(zewaut(rr)*(1-lamdat)+(lamdat*Cus(rr)));
end;
medzta=mean(Cus);

%Límites de control

LSCewmaa=medzta+ceat*desvaut*sqrt(lamdat/(2-lamdat));
LICewmaa=medzta-ceat*desvaut*sqrt(lamdat/(2-lamdat));
LCewmaa=medzta;

%Vectores a graficar

LimSupEA(1:mon)=LSCewmaa;
LimInfEA(1:mon)=LICewmaa;
LimCentEA(1:mon)=LCewmaa;
zewaut;
```



```
grafewa=input('DIGA true para mostrar el gráfico de control EWMA
AUTOCORRELACIONADO: ')
if grafewa==true

    hold on
    plot(LimSupEA, '-b', 'LineWidth', 1, ...
        'MarkerEdgeColor', 'b', ...
        'MarkerFaceColor', 'b', ...
        'MarkerSize', 5)

    plot(LimCentEA, '-r', 'LineWidth', 1, ...
        'MarkerEdgeColor', 'r', ...
        'MarkerFaceColor', 'r', ...
        'MarkerSize', 5)

    plot(LimInfEA, '-b', 'LineWidth', 1, ...
        'MarkerEdgeColor', 'b', ...
        'MarkerFaceColor', 'b', ...
        'MarkerSize', 5)

    plot(zewaut, '-mo', 'LineWidth', 1, ...
        'MarkerEdgeColor', 'm', ...
        'MarkerFaceColor', 'm', ...
        'MarkerSize', 5)

    xlabel('Tiempo')
    ylabel('Ewma ')
    title('\fontname{andalus}EWMA PARA DATOS
AUTOCORRELACIONADOS', 'FontSize', 14)

    legend('Límite Superior', 'Límite Central', 'Límite
Inferior', 'Z', 0)
    hold off
end;

%CALCULO DE Shi Y DE LA DESVIACION ESTANDAR

miuc(1:1)=mean(dema);
for i=1:ValorN;
    miuc(i+1)=(1-Phi1)*mean(dema)+Phi1*miuc(i);
end;
varu=cov(miuc);
varalfa=varu*(1-Phi1);
varx=((varu)+sqrt(varerror));
desvx=sqrt(varx);
```



```
Shil=((varu)/(varx));
```

```
function [ G ] = Gkn( x, k, n )
%Programa para determinar Gkn
G=((k*Logvik1(x,k,n) +(n-k)*Logvik2(x,k,n)/C_Bartlett(k,n)));
end
```

```
function [ Gmaxn ] = GMAXN( x,n )
%Gmax = Determinacion de Gmaxn
N=n;
Gmaxn=zeros(N,1);
for z=10:N;
    Gmax=0;
    for k=2:(z-2);
        GKN=Gkn( x, k, z );
        if GKN>Gmax;
            Gmax=GKN;
        end;
    end;
    Gmaxn(z)=Gmax;
end;
warning('off','MATLAB:dispatcher:InexactCaseMatch')
end
```

```
function graficos_aed(x)
% Conjunto de gráficos que apoyan el Análisis de los datos
%
% Gráfico de los datos en su forma original
figure('NumberTitle','off','Name','Datos Originales');

% Create plot
plot(x,'MarkerFaceColor',[0 0 0],'MarkerEdgeColor',[0 0 0],
'MarkerSize',5,...
'Marker','o',...
'LineWidth',1,...
'Color',[1 0 0]);
xlabel('tiempo','FontSize',14,'FontName','Andalus');
ylabel('Datos Originales','FontSize',14,'FontName','Andalus');
title('\fontname{arial}Datos Originales','FontSize',14,'FontName','Andalus');
datacursormode on
hold off
```



```
% Gráficos Preliminares
%
% Histograma
figure('Name','Gráficos Preliminares','NumberTitle','off');
hold on
subplot(2,2,1:2);
histfit(x)
    xlabel('Datos Originales')
    ylabel('Frecuencia')
    title('\fontname{arial}Histograma','FontSize',14)
    datacursormode on

%
% Diagrama de Caja
subplot(2,2,3);
boxplot(x)
    ylabel('Datos Originales')
    title('\fontname{arial}Gráfico de Caja y Bigotes','FontSize',14)
    datacursormode on

%
% Gráfico Probabilístico Normal
subplot(2,2,4);
probplot(x)
    title('\fontname{arial}Gráfico de Probabilidad Normal','FontSize',14)
    datacursormode on
    hold off

end

function graficos_aed_box(x)
% Conjunto de gráficos que apoyan el Análisis de los datos
% después de la transformación BOX-COX
%
% Gráficos después de la transformación BOX-COX
%
% Histograma
figure('Name',' Datos Box-Cox ','NumberTitle','off');
hold on
subplot(2,2,1:2);
histfit(x)
    xlabel('Datos Transformados','FontSize',14,'FontName','Andalus')
    ylabel('Frecuencia','FontSize',14,'FontName','Andalus')
    title('\fontname{andalus}\fontsize{20}Histograma')
    datacursormode on

%
% Diagrama de Caja
subplot(2,2,3);
boxplot(x)
    ylabel('Datos Transformados','FontSize',12,'FontName','Andalus')
    title('\fontname{andalus}\fontsize{20}Gráfico de Caja y Bigotes')
```



```
        datacursormode on
%
% Gráfico Probabilístico Normal
subplot(2,2,4);
probplot(x)
    title('\fontname{andalus}\fontsize{20}Gráfico de Probabilidad
Normal')
    xlabel('Data', 'FontSize', 12, 'FontName', 'Andalus')
    ylabel('Probabilidad', 'FontSize', 12, 'FontName', 'Andalus')
        datacursormode on
        hold off
end

function graficos_aed_dif(x)
% Conjunto de gráficos que apoyan el Análisis de los datos
% después de las diferenciaciones
%
% Gráficos despues de la diferenciaciones en la media
%
% Histograma
figure('Name', ' Datos Diferenciaciones ', 'NumberTitle', 'off');
hold on
subplot(2,2,1:2);
histfit(x)
    xlabel('Datos Diferenciados', 'FontSize', 14, 'FontName', 'Andalus')
    ylabel('Frecuencia', 'FontSize', 14, 'FontName', 'Andalus')
    title('\fontname{andalus}\fontsize{20}Histograma')
        datacursormode on
%
% Diagrama de Caja
subplot(2,2,3);
boxplot(x)
    ylabel('Datos Diferenciados', 'FontSize', 12, 'FontName', 'Andalus')
    title('\fontname{andalus}\fontsize{20}Gráfico de Caja y Bigotes')
        datacursormode on
%
% Gráfico Probabilístico Normal
subplot(2,2,4);
probplot(x)
    title('\fontname{andalus}\fontsize{20}Gráfico de Probabilidad
Normal')
    xlabel('Data', 'FontSize', 12, 'FontName', 'Andalus')
    ylabel('Probabilidad', 'FontSize', 12, 'FontName', 'Andalus')
        datacursormode on
        hold off
end
```





```
function graficos_aed_dep(x)
% Conjunto de gráficos que apoyan el Análisis de los datos
% despues que la muestra ha sido depurada
%
% Gráfico de los datos en su forma depurada
figure('NumberTitle','off','Name','Datos Depurados');

% Create plot
plot(x,'MarkerFaceColor',[0 0 0],'MarkerEdgeColor',[0 0 0],
'MarkerSize',5,...
'Marker','o',...
'LineWidth',1,...
'Color',[1 0 0]);
xlabel('tiempo','FontSize',14,'FontName','Andalus');
ylabel('Datos Depurados','FontSize',14,'FontName','Andalus');
title('\fontname{arial}Datos
Depurados','FontSize',14,'FontName','Andalus');
datacursormode on
hold off

% Gráficos Preliminares
%
% Histograma
figure('Name','Gráficos Preliminares 2','NumberTitle','off');
hold on
subplot(2,2,1:2);
histfit(x)
xlabel('Datos Depurados')
ylabel('Frecuencia')
title('\fontname{arial}Histograma','FontSize',14)
datacursormode on

%
% Diagrama de Caja
subplot(2,2,3);
boxplot(x)
ylabel('Datos Depurados')
title('\fontname{arial}Gráfico de Caja y Bigotes','FontSize',14)
datacursormode on

%
% Gráfico Probabilístico Normal
subplot(2,2,4);
probplot(x)
title('\fontname{arial}Gráfico de Probabilidad
Normal','FontSize',14)
datacursormode on
hold off

end
```



```
function [ meanX ] = Xik( x, i, k )
%Xik media entre de i a k del vector x
suma=0;
for n=(i+1):k
    suma=suma+x(n);
end;
meanX=suma/(k-i);
warning('off', 'MATLAB:dispatcher:InexactCaseMatch')
end

function [ VARX ] = Vik( x,i,k )
%Vik Varianza entre i y k del vector x
suma=0;
for n=(i+1):k;
    suma=suma+((x(n)-Xik(x,i,k))^2);
end;
VARX=suma;
end

function [solt] = taprox( X,k,n )
%UNTITLED Summary of this function goes here
% Detailed explanation goes here
t_aprox=(Xik(X,0,k)-Xik(X,k,n))/(sqrt(((Vik(X,0,k))/(k-1))
+((Vik(X,k,n))/(n-k-1))));
r=((((Vik(X,0,k))/(k-1))+((Vik(X,k,n))/(n-k-1)))^2)/
(((Vik(X,0,k))/(k-1))^2+((Vik(X,k,n))/(n-k-1))^2));
t=tstat(r);
if (t_aprox>=t)
    solt=true;
else
    solt=false;
end;
end

function [ TablaH ] = TabH(n,alfa)
%Lectura de la tabla del estadistico H desarrollado por Hokin
%y calculo del mismo a partir del dato 14
Z=n;
TablaH=zeros(Z,1, 'double');
for N=(10:Z);
    if N>14
        if 0.001<=alfa
            if (alfa<0.05)
```



```
        TablaH(N)=1.58-
2.52*log(alfa)+(0.094+.33*log(alfa)/(sqrt(N-9)));
        end;
    end;
    if (alfa==0.05)
        TablaH(N)=8.43+.074*log(N-9);
    end;
else
    TabAPC=xlsread('TabHawkins');
    [nAPC,pAPC]=size(TabAPC);
    for i=2:nAPC;
        ValAPC=TabAPC(i,1);
        if ValAPC==N;
            for j=2:pAPC;
                Val2APC=TabAPC(1,j);
                if Val2APC==alfa;
                    TablaH(N)=TabAPC(i,j);
                end;
            end;
        end;
    end;
end;
end;
end

function [yy]=qui_neg(xx)
%Cambio de Negativo a Positivo, Si el vector tiene Negativos
[n]=size(xx);
cc=max(n);
bandn=false;
menv=10000000;
for i=1:cc;
    if xx(i)<0 && bandn==false;
        bandn=true;
        menv=min(xx);
    end;
end;
% formación de vector positivo
if bandn==true;
    gg=zeros([cc 1]);
    for i=1:cc;
        gg(i,1)=xx(i)-menv+0.001;
    end;
    yy=gg;
end
```



```
function [punto_de_cambio]=Punto_de_Cambio(x,CAMB)
%programa para determinar el punto donde se estima un cambio
Gmax=0;
for k=2:(CAMB-2);
    GKN=Gkn( x, k, CAMB );
    if GKN>Gmax;
        Gmax=GKN;
        punto_de_cambio=k;
    end;
end;
warning('off','MATLAB:dispatcher:InexactCaseMatch')
end

dema=vector_original;
%Determinación de los parámetros del modelo
if ESAR==true;
    [a,b,c]=arburg(dema,1);
    varerror=b;
    Phil=c;
end;
if ESARMA==true;
    dema'
    disp('A continuación usted deberá interactuar con el programa
MINITAB')
    disp('Asegúrese de abrir dicho programa')
    disp('Debe transferir la data DEMA a una columna en MINITAB')
    disp('Seleccione la data que esta en pantalla')
    autorizar11=input('Realizó la instrucción? Diga "true" Si ya lo
hizo: ')
    if autorizar11==true
        disp('Cópiela en una columna en MINITAB')
        autorizar12=input('Realizó la instrucción? Diga "true" Si ya lo
hizo: ')
        if autorizar12==true
            disp('En el panel de tareas debe seleccionar el ícono
"STAT"')
            disp('Dentro de "STAT" seleccione el ícono "TIME SERIES"')
            disp('Dentro de "TIME SERIES" seleccione el ícono "ARIMA"')
            disp('Se abrirá una ventana')
            autorizar13=input('Realizó la instrucción? Diga "true" si ya
lo hizo: ')
            if autorizar13==true
                disp('En el espacio denominado "SERIES" coloque el nombre
de')
                disp('la columna donde se encuentra la data DEMA')
                disp('En el espacio "AUTOREGRESSIVE" coloque el valor de
"1"')
```



```
disp('En el espacio "MOVING AVERAGE" coloque el valor de
"1"')
autorizar14=input('Realizó la instrucción? Diga "true" si
ya lo hizo: ')
if autorizar14==true
    disp('Verifique y luego presione "OK"')
    disp('Se cerrará la ventana y aparecerá la principal
de MINITAB')
    disp('dividida en dos partes, una de ellas dice
"SESSION".')
    disp('Con el botón izquierdo del mouse haga click
dentro de esa pantalla,')
    disp('suba a través de ella hasta que consiga un
título que diga')
    disp('lo siguiente: ')
    disp('Final Estimates of Parameters')
    disp('Type          Coef  SE Coef    T      P')
    disp('AR   1          0,1979  0,3730    0,53  0,597')
    disp('MA   1          0,4422  0,3410    1,30  0,198')
    disp('Constant -0,13078  0,05288  -2,47  0,015')
    disp('Mean          -0,16304  0,06593')
    disp('Number of observations: 100')
    disp('Residuals:   SS = 87,1556 (backforecasts
excluded)')
    disp('MS = 0,8985  DF = 97')
    disp('Esto es un ejemplo ilustrativo para guiarlo')
    autorizar15=input('Realizó la instrucción? Diga "true"
si ya lo hizo: ')
    if autorizar15==true
        disp('Verifique el valor "COEF" correspondiente
al "AR 1"')
        ar1=input('Coloque ese valor aqui: ')
        disp('Verifique el valor "COEF" correspondiente
al "MA 1"')
        ma1=input('Coloque ese valor aqui: ')
        disp('Seleccione y copie el valor que esta al
lado de "Number of observations"')
        nobs=input('Pegue el valor ahora: ')
        end;
        InfiniteAR = garchar(ar1,ma1,nobs);
        dema=InfiniteAR;
        [aa,bb,cc]=arburg(dema,1);
        varerror=bb;
        Phil=cc;
    end;
end;
end;
end;
end;
```



```
function index = outliers( y, alpha, k)
% Detección y eliminación de outliers en la muestra
% Test de ROSNER
y = y(:);
n = length( y );
r = zeros( k, 1 );

%% calculate statistics for up to k outliers
for i = 1:k
ys = abs(y-mean(y))/std(y);
posmax=1;
for a=2:length(y)
    if ys(a)>ys(posmax)
        posmax=a;
    end
end
r(i)=ys(posmax,1);
yyy=[];
m=0;
for l= 1:length(y)
    if l~=posmax
        m=m+1;
        yyy(m,1)=y(l);
    end
end
y = [];
for l = 1 :length(yyy)
y(l,1) = yyy(l,1);
end;
end
%% Prueba estadística para encontrar outliers
for i = 1:k
pcrit=1-(alpha/((2*(n-i+1))));
t=tinv(pcrit, n-i-1);
lambda(i,1)=(n-i)*t./sqrt(((n-i-1+t^2)*(n-i+1)));
end;
%Reporte de resultados
tv=0;
for w=1:k
if r(w) > lambda(w)
    tv=w;
end
end
disp('Outliers detectados=')
disp(tv)
index=tv;
end
```



```
function [yy,cus]= matriz_vector(xx)
%creacion del vector si p es mayor que 1
[n,p]=size(xx);
cc=max(n);
dd=max(p);
vectoru=zeros([cc 1]);
for i=1:cc;
    sumav=0;
    for j=1:dd;
        sumav=sumav+xx(i,j);
    end;
    sumav1=sumav/dd;
    vectoru(i,1)=sumav1;
end;
yy=vectoru;
cus=vectoru;
```

```
function [ LOG2 ] = Logvik2( x, k, n )
%programa para determinar el log de la varianza de 0 a n entre la
varianza
%de n a k
LOG2=log((Vik(x,0,n)/(n-0))/(Vik(x,k,n)/(n-k)));
end
```

```
function [ LOG1 ] = Logvik1( x, k, n )
%programa para determinar el log de la varianza de 0 a n entre la
varianza
%de 0 a k
LOG1=log((Vik(x,0,n)/(n-0))/(Vik(x,0,k)/(k-0)));
end
```

```
function [n, p, yy,Cus]= lec_dat_orig
%Lectura de Datos Originales
disp('A continuación debe leer el archivo que contiene los datos a
estudiar')
disp('para leerlos debe colocar el nombre del archivo EXCEL que
contiene')
disp('los datos a estudiar, hagalo cuando se le indique')
filename=input('Indique el nombre del archivo a leer: ','s');
yy=xlsread(filename);
[n,p]=size(yy);
Cus=yy;
```



```
%Lectura de data de tabla de EWMA independiente
```

```
disp('A continuación deberá leer la tabla que contiene los valores de  
LAMDA Y K')  
disp('para leer el archivo correcto debe colocar el nombre "EWMA_I"  
cuando')  
disp('se le indique, (Debe cuidar usar mayúsculas y el guión bajo)')
```

```
filename = input('Indique el nombre de archivo a leer: ', 's')  
ValEwmaI = xlsread(filename);  
[newi,pewi]=size(ValEwmaI);
```

```
%Lectura de data de tabla de EWMA autocorrelacionado  
%para obtener el valor de C
```

```
disp('A continuación deberá leer el archivo EXCEL 2003 que contiene los  
valores')  
disp('de C en función de Phi, Shi y de LAMDA')  
disp('para leer el archivo deberá colocar el nombre "EWMA_AC" cuando  
se le indique')  
disp('cuide colocar el nombre en mayúsculas y colocar el guión bajo')
```

```
filename = input('Indique el nombre de archivo a leer: ', 's')  
VEC = xlsread(filename);  
[neac,peac]=size(VEC);
```

```
%Lectura de data de tabla de EWMA autocorrelacionado  
%para obtener el valor de LAMDA
```

```
disp('A continuación deberá leer el archivo EXCEL 2003 que contiene los  
valores')  
disp('de LAMDA en función de Phi, Shi y el cambio en la media')  
disp('para leer el archivo deberá colocar el nombre "EWMA_A" cuando se  
le indique')  
disp('cuide colocar el nombre en mayúscula y colocar el guión bajo')
```

```
filename = input('Indique el nombre de archivo a leer: ', 's')  
ValEwmaA = xlsread(filename);  
[newa,pewa]=size(ValEwmaA);
```





```
%Lectura de la data de tabla de CUSUM autocorrelacionado
%para obtener el valor de C

disp('A continuación deberá leer el archivo EXCEL 2003 que contiene los
valores')
disp('de LAMDA en función de Phi, Shi y el cambio en la media')
disp('para leer el archivo deberá colocar el nombre "CUSUM_AC" cuando
se le indique')
disp('cuide colocar el nombre en mayúsculas y colocar el guión bajo')

filename = input('Indique el nombre de archivo a leer: ', 's')
VECC = xlsread(filename);
[neacc,peacc]=size(VECC);

%Lectura de data de tabla de CUSUM autocorrelacionado
%para obtener el valor de r

disp('A continuación deberá leer el archivo EXCEL 2003 que contiene los
valores')
disp('de LAMDA en función de Phi, Shi y el cambio en la media')
disp('para leer el archivo deberá colocar el nombre "CUSUM_AU" cuando
se le indique')
disp('cuide colocar el nombre en mayúsculas y colocar el guión bajo')

filename = input('Indique el nombre de archivo a leer: ', 's')
ValCusumA = xlsread(filename);
[ncua,pcua]=size(ValCusumA);

function [ h ] = H1( n,alfa )
%UNTITLED2 Summary of this function goes here
% Detailed explanation goes here
h=1.58-2.52*log(alfa)+(0.094+.33*log(alfa)/(sqrt(n-9)));
end
```



## BIBLIOGRAFÍA

APARISI, F. y GARCÍA-DÍAZ, J. C. (2004). “Optimization of univariate and multivariate exponentially weighted moving average control charts using genetic algorithms”. *Computers and Operations Research*, 31, pp. 1437–1454.

ARRANZ, J. M. y ZAMORA, M. M. (2000). “Análisis de autocorrelación”.

BARNETT, V., & LEWIS, T. (1978). *Outliers in statistical data*. New York: Wiley.

BECKMAN, R. J., AND COOK, R. D. (1983). “Outlie.....s”. *Technometrics*, 25, pp. 119–149.

BORROR, C. M., MONTGOMERY, D. C. y RUNGER G. C. (1999). “Robustness of the EWMA Control Chart to Non-normality”. *Journal of Quality Technology*, 31, pp. 309–316.

BOX, G. y COX, D. (1964). “An Analysis of Transformations”. *Journal of the Royal Statistical Society*, Series B:211–264.

BOX, G. E. P., JENKINS, G. M. y REINSEL, G. C. (1994). “*Time Series Análisis*”. 3ra. Edición. New Jersey, Estados Unidos.



BROCKWELL, P. J. AND DAVIS, R. A. (1987). *Time Series Analysis: Theory and Methods*. Springer - Verlag, New York, NY.

CARRASQUEL, A., GONZÁLEZ, Y. y RODRÍGUEZ, B. (2001). “*Análisis Econométrico con EViews* ”. 1º Edición. Editorial Ctra. De canilas, Madrid 144.

CHATFIELD, C. (2003). “*The Analysis of Time Series: An Introduction*”. 6th edition. Boca Raton: Chapman & Hall/CRC, New York.

CROWDER, S. V. (1987). “Average lengths of exponentially weighted moving average control charts”. *Journal of Quality Technology*, 19, pp 161–164.

CROWDER, S. V. (1989). “Design of Exponentially Weighted Moving Average Schemes”. *Journal of Quality Technology*, 21, pp. 155–162.

CROMWELL, J., LABYS, W., TERRAZA, M. (1994). “Univariate Tests for time series models”. *Sage University Paper Series on Quantitative Applications in the social sciences*. Series N°07–099.

DICKEY, D. y FULLER, W. (1979). “Distribution of the Estimators for Autoregressive Time Series with a Unit Root”. *Journal of the American Statistical Association*, 74, pp. 427–31.



DURBIN, J. y WATSON, G. S. (1950; 1951; 1971). “Testing for serial correlation in least-squares regression”. I, II and III, *Biometrika*, 37, pp. 409–428; 38, pp. 159–178; 58, pp. 1–19.

EYSSAUTIER, M. (2006). “*Metodología de la Investigación: Desarrollo de la Inteligencia*”. 5ta. Edición. Editorial International Thomson, México.

FAREBROTHER R.W. (1980). “The Durbin–Watson Test for Serial Correlation when there is No Intercept in the Regression”. *Econometrica*, 48, pp. 1553–1563.

FERRER, A. (2004). “Control estadístico de procesos con dinámica: revisión del estado del arte y perspectiva de futuro”. *Estadística Española*. 46, pp. 19–47.

GEL, Y. (2009). “*Introduction to Change Point Analysis*”. Proceedings del VI Encuentro Colombia – Venezuela de Estadística y VIII Jornada de Aplicaciones Matemáticas, Universidad de Carabobo Valencia, Venezuela.

GUJARATI, D. (2004), “*Basic Econometrics*”. 4th Edition. The McGraw–Hill Companies.



GUTIÉRREZ, P. H. y DE LA VARA, S. R. (2004). “*Control Estadístico de Calidad y Seis Sigma*”. McGRAW-HILL Interamericana Editores, Mexico.

HANKE, J. y REITSCH, A. (2000), “*Pronóstico de los negocios*”. 6ta. Edición. México.

HAWKINS, D. M. (1980). “*Identification of outliers*”. Chapman and Hall, New York, NY.

HAWKINS, D. M., y OLWELL, D. H. (1997). “*Cumulative Sum Chart and Charting for Quality Improvement*”. Editorial Springer, Verlag New York.

HAWKINS, D. M., y ZAMBA, K. D. (2005a). “Change Point Model for Statistical Process Control with Shift in Variance”. *Journal of Quality Technology*, 37, pp. 21–31.

HAWKINS, D. M., y ZAMBA, K. D. (2005b). “Statistical Process Control for Shift in Mean or Variance Using the Change Point Formulation”. *Technometrics*, 47, pp. 164–173.

HERNÁNDEZ, A. J. (2007). “*Análisis de Series de Temporales Económicas II*”. 1ra edición. Editorial Esic, Madrid España.



HERNANDEZ, S.(2006).”*Metodología de la investigación*”.Editorial Mc Graw-Hill”.3da edición. México.

KENDALL M. G. AND. BUCKLAND W. R. (1971). “*A Dictionary of Statistical Terms*”. Hafner Publishing Company, New York.

HUNTER, J. S. (1986). “The Exponentially Weighted Moving Average”. *Journal of Quality Technology*, 18, (4), pp. 203–210.

LAWLEY D. N. (1956). “A general method of approximating to the distribution of likelihood ratio criteria”. *Biometrika*. 43, pp. 295–303.

LU, C.W. AND REYNOLDS, JR. M. R. (1999) “EWMA Control Charts for Monitoring the Mean of Autocorrelated Processes” *Journal of Quality Technology*. 31, pp. 166–188.

LU, C.W. AND REYNOLDS, JR. M. R. (2001) “Cusum Charts For Monitoring  
An Autocorrelated Process” *Journal of Quality Technology*. 33 pp. 316–334.

MARKS, R.G., AND RAO, P.V. (1979) “*An estimation procedure for data containing outliers with a one directional shift in the mean*” J. Amer. Statist. Assoc. 74, 614–620.



MASON, R. L. y YOUNG, J. C. (2002). “*Multivariate statistical process control with industrial applications*”. Editorial Board.

MOHD, N. y BEE, Y. (2011).” Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling test”. *Journal of Statistical modeling and Analytics*. 2, pp. 21-33.

MONTGOMERY, D. C. (2005). “*Introduction to Statistical Quality Control*”. 5th Ed. John Wiley and Sons, New York, NY.

PAGE, E. S. (1954). “Continuous Inspection Schemes”. *Biometrika*, 41, pp. 100-115.

PEÑA, N. y QUIROZ, A. J. (2009). “*Estudio de la robustez y otras propiedades de los métodos gráficos de Control Estadístico de Procesos*”. Proceedings del VI Encuentro Colombia - Venezuela de Estadística y VIII Jornada de Aplicaciones Matemáticas, Universidad de Carabobo Valencia, Venezuela.

PÉREZ, F. O. (2007). “*Introducción a las Series de Tiempo: Métodos paramétricos*”. 1ª edición. Universidad de Medellín, Colombia.



PHILLIPS, P. y PERRON, P. (1988). “Testing for a Unit Root in Time Series Regression”. *Biometrika*. 75, pp. 335-46.

REYNOLDS, M. R., JR.; ARNOLD, J. C.; AND BAIK, J. W. (1996), “Variable Sampling Interval X Chart in the presence of correlation”. *Journal of quality technology*. 28, pp. 12-30.

ROBERTS, S. W. (1959). “Control Charts Based on Geometric Moving Averages”. *Technometrics*, 1, pp. 239-250.

ROSNER, B. (1975), “On the Detection of Many Outliers”. *Technometrics*, 17, pp. 221-227.

SHEWHART, W. (1931). “*Economic Control of Quality of Manufactured Product*”, D. Van Nostrand Co, Inc. New York, NY.

TAMAYO Y TAMAYO, M. (2004). “*El Proceso de la Investigación Científica*”. 4ta. Edición. Limusa, México.

TAYLOR, W. (2000a). “Change-point Analysis: A Powerful New Tool for Detecting Changes”. <http://www.variation.com/cpa/tech/changepoint.html>.





TAYLOR, W.(2000b). “A Pattern Test for Distinguishing Between Autoregressive and Mean-shifted Data”. <http://www.variation.com/cpa/tech/pattern.html>.

TUKEY, J. W. (1977). “*Exploratory Data Analysis*”. Addison-Wesley Publishing Company.

WALPOLE, R. E. (1999). “*Probabilidad y Estadística para Ingenieros*”. 6a Edición. Prentice – Hall Hispanoamericana, S.A.. México.

American Statistical Association (2012). [Página Web en línea]. Revisado el 15 de mayo de 2012. Disponible en: [www.amstat.org](http://www.amstat.org)



---

## CONCLUSIONES

Basados en los objetivos propuestos y en los resultados obtenidos, se concluye lo siguiente:

- El programa propuesto realiza el análisis e interpretación de datos, a través de diferentes procesos estadísticos y econométricos. En el caso de las series de tiempo permite una visualización de comportamiento de los datos antes y después de las distintas transformaciones aplicadas a la muestra.
- Al observar el contraste desde la detección de los valores atípicos hasta la eliminación de los mismos de los ejemplos realizados, se puede concluir que la presencia de valores atípicos influye considerablemente en los resultados obtenidos, debido a que éstos alteran el comportamiento de la data.
- El programa diseñado en este trabajo, permite al usuario elegir trabajar con una data original o depurada; en tal caso si existe la presencia de valores atípicos, este decide si representan una observación discordante, contaminante o influyente y su exclusión o no del estudio.
- El programa diseñado permite elegir el nivel de confiabilidad y la cantidad de posibles valores atípicos presentes en la muestra.



- Al trabajar en la construcción de los gráficos de control se demuestra que la etapa de análisis de datos disminuye posibles falsas alarmas que se generan en los gráficos de control.
- El programa diseñado logra el objetivo de desarrollar una herramienta computarizada para el análisis de datos y su adaptación para la construcción de los gráficos de control EWMA, CUSUM y APC, otorgándole al analista una herramienta que le facilite el trabajo en el Control de Calidad.



## RECOMENDACIONES.

A continuación se sugieren una serie de recomendaciones referentes a los aspectos planteados en el marco de esta investigación:

- A fin de complementar las funciones del programa planteado en este trabajo especial de grado, se sugiere incluir un estudio que permita analizar las series de tiempo que contenga la dependencia específicamente no-lineal e incluir modelos para la media y la varianza.
- Es importante profundizar en el estudio acerca de la eliminación de estacionariedad, ya que en este trabajo se llegó hasta la aplicación de la segunda diferencia para corregir este efecto, existiendo procedimientos a aplicar, luego de la segunda diferencia, que se pueden incluir en este estudio posteriormente.
- Se recomienda seguir con la investigación acerca de la teoría existente en el contraste  $h$  de Durbin, para incluir el procedimiento que se debe aplicar cuando el radicando es negativo, para que esta prueba pueda ser usada cuando esto ocurra.
- Sería de gran aporte continuar con el estudio comenzado en este trabajo especial de grado, debido a la gran utilidad que tiene en el mejoramiento



del control de la calidad en los procesos industriales actuales, que a medida que pasa el tiempo reclaman mayor precisión en la determinación de cambios en el proceso.

- Se recomienda la inclusión del estudio de series de tiempo como materia electiva, ya que, la mayoría de la data de los procesos están en función del tiempo, por lo tanto debe ser tratada como tal.
- Se sugiere incluir un sub-programa que realice la eliminación de la autocorrelación para los procesos donde este presente.



**TABLA DE CONTENIDO**

Agradecimientos.....  
.....i

Dedicatoria.....  
.....iii

RESUMEN.....  
.....viii

INTRODUCCIÓN ..... 1

CAPITULO I. EL PROBLEMA..... 5

    Planteamiento del problema ..... 5

    Objetivos ..... 8

        Objetivo general ..... 8

        Objetivos específicos ..... 8

    Alcance de la investigación ..... 9

    Justificación..... 10

CAPITULO II. MARCO TEÓRICO ..... 13

    Bases Teóricas..... 13

        Naturaleza y fuentes de información para el análisis de  
datos.....13

        Series de Tiempo ..... 13

        Análisis Exploratorio de  
datos.....15

        Procesos Estocásticos ..... 24



Procesos Estacionarios.....	26
Estacionariedad .....	28
Pruebas de Estacionariedad .....	28
Prueba de Raíz Unitaria de Dickey-Fuller .....	29
Prueba de Raíz Unitaria de Phillips- Perron.....	31
Transformación para Obtener Series de Tiempo Estacionarias .....	31
Corrección de la no Estacionariedad en la Varianza.....	32
Corrección de la no Estacionariedad en la Media .....	33
Método de Box & Jenkins.....	34
Modelo Promedio Móvil (MA(q) ).....	37
Modelo Autorregresivos (AR) .....	38
Modelo ARMA(p,q) .....	40
Autocorrelación .....	42
Detección de Autocorrelación.....	45
Contraste de $d$ de Durbin-Watson.....	47
Estadístico $h$ de Durbin .....	53
Gráficos de Control.....	56
Grafico de control Shewhart.....	56



Gráficos de Control CUSUM .....	60
Gráficos de Control EWMA.....	63
Análisis de Punto de Cambio.....	67
Modelo Bayesiano para Análisis de Punto de Cambio.....	68
Esquema del gráfico de control CUSUM para datos autocorrelacionados.....	78
Esquema del gráfico de control EWMA para datos autocorrelacionados.....	82
Comportamiento de los datos en los procesos productivos.....	85
Lenguaje de Programación MATLAB .....	86
CAPITULO III. METODOLOGÍA .....	91
Construcción del modelo y pruebas para la serie de tiempo.....	91
Metodología propuesta .....	91
CAPITULO IV. DISEÑO DE LA HERRAMIENTA.....	111
Entrada de base de datos.....	113
Análisis preliminares.....	1
13	
Análisis Exploratorio de Datos.....	114





---

Estudio de estacionariedad.....	116
Estudio de autocorrelación .....	118
Metodología Box- Jenkins.....	119
Gráficos de control EWMA para datos no autocorrelacionados.....	121
Gráficos de control CUSUM para datos no autocorrelacionados .....	122
Gráficos de control EWMA para datos autocorrelacionados .....	123
Gráficos de control CUSUM para datos autocorrelacionados .....	124
Análisis de punto de cambio .....	125
CAPITULO V. APLICACIÓN DE LA HERRAMIENTA .....	129
Ejemplo N° 1. ....	129
Ejemplo N° 2.....	151
Ejemplo N° 3.....	160
Ejemplo N° 4.....	167
CONCLUSIONES .....	181
RECOMENDACIONES.....	185
APENDICE I.....	189
APENDICE II .....	235
BIBLIOGRAFÍA.....	251

