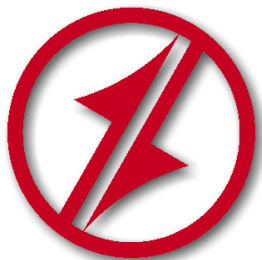
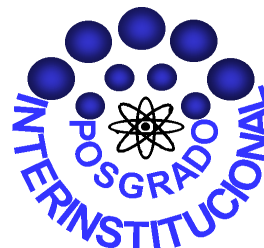


CORPORACIÓN MEXICANA DE INVESTIGACIÓN EN MATERIALES

DIVISION DE ESTUDIOS DE POSGRADO



COMIMSA



**CIENCIA Y
TECNOLOGIA**

Gráficos de control para datos serialmente correlacionados: caso Sanden Mexicana.

TESIS

Que para obtener el Grado Académico de

Maestro en Ciencia y Tecnología en la Especialidad de Ingeniería Industrial y de Manufactura

**Presenta:
Marisol de la Garza Cervantes.**

Saltillo, Coah., Diciembre 2011

Gráficos para datos serialmente correlacionados. Caso Sanden Mexicana.

Por
Marisol de la Garza Cervantes

Tesis

Presentada al Programa Interinstitucional en Ciencia y tecnología

Sede

Corporación Mexicana de Investigación en Materiales, S.A. de C.V.

Como requisito parcial para obtener el Grado Académico de

Maestro en Ciencia y Tecnología
Especialidad en Ingeniería Industrial y de Manufactura

Programa Interinstitucional en Ciencia y Tecnología COMIMSA / CONACyT

Saltillo, Coah., Diciembre 2011

Corporación Mexicana de Investigación en Materiales, S.A de C.V
Gerencia de Desarrollo Humano
División de Estudios de Posgrado

Los abajo firmantes, miembros del Comité Tutorial recomendamos que la Tesis
**“GRÁFICOS DE CONTROL PARA DATOS SERIALMENTE
CORRELACIONADOS. CASO SANDEN MEXICANA.”**, realizada por el
alumno **MARISOL DE LA GARZA CERVANTES** matrícula **0906IM3058** sea
aceptada para su defensa como Maestro en Ciencia y Tecnología con
Especialidad en **Ingeniería Industrial y de Manufactura**.

El Comité Tutorial

Tutor Académico

Director de Tesis

Dr. Mario Cantú Sifuentes.

Tutor en Planta

Ing. Fidel López Vallejo

Vo. Bo

M.C Claudia A. González Rodríguez
Coordinación General de Estudios de Posgrado
COMIMSA

Corporación Mexicana de Investigación en Materiales
Gerencia de Desarrollo Humano
División de Estudios de Posgrado

Los abajo firmantes, miembros del Jurado del Examen de Grado del alumno, **MARISOL DE LA GARZA CERVANTES** una vez leída y revisada la tesis titulada “**GRÁFICOS DE CONTROL PARA DATOS SERIALMENTE CORRELACIONADOS. CASO SANDEN MEXICANA.**”, aceptamos que la referida tesis revisada y corregida sea presentada por el alumno para aspirar al grado de Maestro en Ciencia y Tecnología con Especialidad en Ingeniería Industrial y de Manufactura durante el Examen de Grado correspondiente.

Y para que así conste firmamos la presente a los (días) del mes de (mes) del año (año).

Dr. Federico Zertuche Luis.
Presidente

Dr. José Antonio Vázquez
López.
Secretario

Dr. Mario Cantú Sifuentes.
Vocal

Dr. Eduardo Marroquín Prado.
Vocal

Dr. David Salvador González
González.
Vocal

DEDICATORIA.

“Y todo lo que hagáis, hacedlo de corazón, como para *Dios*”.

Colosenses 3.23a.

A mi familia y amigos:

Vivir con propósito es la única manera de vivir de verdad.

Hechos 13:36.

AGRADECIMIENTOS.

Comunicación y apoyo por parte del comité tutorial, vinculación con el sector productivo y dedicación del estudiante son aspectos claves para la culminación de cualquier proyecto.

Agradezco a CONACYT y COMIMSA por los apoyos recibidos en cuanto a becas de manutención, transporte, asistencia a congreso y seminarios; se refiere. A mi comité tutorial por las enseñanzas impartidas en clases, el apoyo brindado para disponer de libros, artículos, material de estudio y asesorías.

También agradezco la disposición de la empresa SANDEN MEXICANA S.A. de C.V. por permitir que el caso de estudio de este proyecto se realizara dentro de sus instalaciones.

A mis compañeros Gladys y Abraham por el apoyo brindado dentro de los laboratorios para realizar las diferentes pruebas mecánicas requeridas en el caso de estudio.

Finalmente, como investigador me comprometo a mejorar mi desempeño como profesionista y motivo a los centros de investigación y empresas a renovar constantemente sus métodos y a buscar una mejora continua.

RESUMEN.

Variación en la calidad de la materia prima, errores en la calibración de un instrumento de medición, falla en algún componente del equipo, desgaste de una herramienta, cambios graduales de presión o temperatura, fatiga, cambios y recesos de operadores, periodicidad en el comportamiento de la máquina por mantenimiento y periodicidad en las propiedades químicas del material son causas que provocan tendencias, patrones cíclicos y cambios en el comportamiento de las mediciones realizadas.

Cuando las observaciones que se grafican en una carta de control no cumplen con el supuesto de independencia, el desempeño de la carta resulta ineficiente. Es decir, cuando los datos del proceso exhiben *autocorrelación* o son *serialmente correlacionados*, se utiliza un modelo autorregresivo de orden uno, $AR(1)$, para explicar su comportamiento. Además es necesario emplear *cartas de control modificadas* que cuenten con la capacidad de tratar este tipo de datos. De acuerdo con *Shumway (2006)*, la *autocorrelación positiva* se presenta en muchas situaciones prácticas, por esto, contar con una metodología de monitorización adecuada y eficiente, permite obtener mejores resultados.

CONTENIDO

Lista de Tablas.....	ix
Lista de Figuras	x
CAPÍTULO 1. INTRODUCCIÓN	12
CAPÍTULO 2. PLANTEAMIENTO DEL PROBLEMA.....	14
2.1 Descripción del Problema.....	14
2.2 Preguntas de Investigación	16
2.3 Hipótesis y Variables de Investigación	16
2.4 Objetivo General.....	17
2.4.1 Objetivos Específicos.....	17
2.5 Justificación	17
2.5.1 Justificación Teórica.....	17
2.5.2 Justificación Práctica.....	18
2.6 Delimitaciones.	18
CAPÍTULO 3. MARCO DE REFERENCIA.....	19
3.1 Marco Teórico.....	19
3.2 Marco Conceptual.....	26
3.2.1 Variación en los Procesos.....	26
3.2.2 Control Estadístico del Proceso: Cartas de control	29
3.2.3 Cartas de Control Clásicas.....	42
3.2.4 Series de Tiempo	47
3.2.5 Medidas de Dependencia: Autocorrelación.....	53
3.2.6 Procesos Estacionarios.....	58
3.2.7 Modelos Autorregresivos AR(p)	63
CAPÍTULO 4. MATERIALES Y MÉTODOS.....	67
CAPÍTULO 5. ANÁLISIS DE RESULTADOS.....	74
Metodología del Proyecto de Investigación.	80
5.1 Etapa 1: Efectos de ignorar la Correlación Serial	80
5.2 Etapa 2: Construcción Teórica de Cartas de Control Clásicas	81

<i>5.3 Etapa 3: Evaluar el ARL para las Cartas Clásicas.....</i>	<i>84</i>
<i>5.4 Etapa 4: Construcción de Cartas Modificadas y su ARL.</i>	<i>94</i>
<i>5.5 Etapa 5: Comparaciones entre Cartas Clásicas y Modificadas.</i>	<i>99</i>
CAPÍTULO 6. CONCLUSIONES Y RECOMENDACIONES.....	103
BIBLIOGRAFÍA.....	107
ANEXOS.....	111
APÉNDICE.....	132

LISTA DE TABLAS

Tabla 5.1 Efectos de ignorar la Correlación Serial Positiva.	80
Tabla 5.2 Efectos de ignorar la Correlación Serial Negativa.....	81
Tabla 5.3 Constantes $d_2(n)$ y $d_3(n)$ para diferentes Tamaños de Muestra.	83
Tabla 5.4 <i>ARL</i> para Carta de Medias Clásica, con diferentes valores de δ y n .	87
Tabla 5.5 <i>ARL</i> para Carta de Medias Clásica, $n=4$	88
Tabla 5.6 <i>ARL</i> para Carta de Medias Clásica, $n=5$	89
Tabla 5.7 <i>ARL</i> para Carta de Rangos Clásica.	91
Tabla 5.8 Valores Críticos de la Distribución ji-cuadrada.....	92
Tabla 5.9 <i>ARL</i> para Carta de Varianza Clásica.	93
Tabla 5.10 <i>ARL</i> para Carta Modificada de Medias, $n=4$	96
Tabla 5.11 <i>ARL</i> para Carta Modificada de Medias, $n=5$	96
Tabla 5.12 Estimación de Varianza para un Proceso <i>AR(1)</i>	98
Tabla 5.13 <i>ARL</i> para Carta Modificada de Varianzas.....	99
Tabla 6.1 Efectos de la Autocorrelación sobre los Límites de Control.....	103

LISTA DE FIGURAS

Fig. 3.1 Pensamiento Estadístico para la Mejora de la Calidad.....	26
Fig. 3.2 Gráficas de Puntos y Dispersión con diferente Desviación Estándar. .	28
Fig. 3.3 Gráficas de Puntos y Dispersión con diferente Media.	28
Fig. 3.4 Comportamiento de $c_4(n)$ para diferentes Tamaños de Muestra.	38
Fig. 3.5 Comportamiento de $d_2(n)$ para diferentes Tamaños de Muestra.....	41
Fig. 3.6 Tendencia Ascendente en una Serie de Tiempo.	48
Fig. 3.7 Variación Cíclica en una Serie de Tiempo.	49
Fig. 3.8 Proceso de Ruido Blanco Gaussiano.	49
Fig. 3.9 Proceso de Promedios Móviles.....	50
Fig. 3.10 Proceso Autorregresivo.....	51
Fig. 3.11 Caminata Aleatoria.....	53
Fig. 3.12 Correlograma para una Serie de Tiempo.....	58
Fig. 3.13 Un Proceso con Media Constante.	60
Fig. 3.14 Un Proceso $AR(1)$	66
Fig. 4.1 Metodología del Proyecto de Investigación.....	70
Fig. 4.2 Caso de Estudio.....	73
Fig. 5.1 Proceso de Fabricación de Rotor Automotriz.....	75
Fig. 5.2 Correlograma de la variable <i>Espesor</i>	76
Fig. 5.3 Gráficas de Dispersión para la variable <i>Espesor</i> y sus retrasos.	77
Fig. 5.4 Correlograma de la variable <i>Altura total</i>	78
Fig. 5.5 Carta de Medias para datos individuales de la variable <i>Espesor</i>	79
Fig. 5.6 Curva de <i>ARL</i> : Carta de Medias, $n=2,3,4$ y 5	99

Fig. 5.7 Comparación Curvas de *ARL*: Carta Clásica y Modificada, $\rho=0.1$ 101

Fig. 5.8 Comparación Curvas de *ARL*: Carta Clásica y Modificada, $\rho=0.2$ 101

Fig. 5.9 Comparación Curvas de *ARL*: Carta Clásica y Modificada, $\rho=0.3$ 102

Fig. 5.10 Comparación Curvas de *ARL*: Carta Clásica y Modificada, $\rho=0.4$... 102

CAPÍTULO 1. INTRODUCCIÓN

El Control Estadístico del Proceso (*CEP*), como lo cita *Wieringa (1999)*, corresponde al conjunto de métodos enfocados a mejorar constantemente la calidad, mediante la reducción de la variación. Para lograr este objetivo, se han utilizado muchas herramientas. La primera de ellas se introdujo en 1930, (Cartas clásicas de Shewhart). Sin embargo, *Psarakis y Papaleonida (2007)*, mencionan que la automatización y el desarrollo continuo de la industria han provocado que las cartas convencionales no sean adecuadas para los procesos actuales. Es por eso que se requiere un rediseño de las mismas, de manera que las nuevas cartas tengan la capacidad de trabajar con datos no normales o dependientes.

La presente tesis, se enfoca en estudiar los efectos de la *dependencia entre las observaciones* en las cartas de control más utilizadas en la industria, tales como medias, rangos y varianzas; en procesos estacionarios explicados mediante un Modelo Autorregresivo, *AR(1)*. Actualmente la industria requiere cartas de control de fácil interpretación y de una construcción sencilla, de manera que su utilización pueda ser implementada rápidamente en los procesos con comportamiento autorregresivo. El caso de estudio de esta tesis se realizó en la empresa Sanden Mexicana S.A. de C.V., en un proceso de fabricación para el rotor automotriz que presenta observaciones autocorrelacionadas. El contenido de la tesis se divide en seis capítulos e información complementaria, tales como anexos y apéndice. En el *capítulo 2* se hace una descripción del problema de investigación, incluyendo objetivos, hipótesis y preguntas de investigación.

En cuanto al *capítulo 3*, se hace una revisión de literatura del tema y posteriormente se introduce al lector a la terminología que es utilizada en el resto de la tesis. Con respecto al *capítulo 4*, se explica la metodología que se siguió desde dos enfoques: el teórico y el caso de estudio. En el *capítulo 5*, se estudian los efectos de ignorar la correlación serial en las cartas clásicas, se explica el rediseño que se hace sobre las mismas para mejorar su desempeño y posteriormente se realiza una comparación en términos de la Longitud Promedio de Corrida (*ARL*). En el *capítulo 6*, se describe la validación de las hipótesis, los resultados obtenidos con respecto a los objetivos planteados, recomendaciones y trabajo futuro.

Finalmente en la sección de *Anexos* se encuentra el código de programación ejecutado en el paquete R para generar las figuras y tablas mostradas en el documento. Así mismo en el *Apéndice* se encuentra información complementaria del tema *Distribución de un estadístico de orden* que se menciona en la sección 3.2.2, en el apartado *Método del rango*.

CAPÍTULO 2. PLANTEAMIENTO DEL PROBLEMA

En este capítulo se hace una descripción del problema, donde el término *autocorrelación* o *correlación serial* se menciona por primera vez en el documento. Además, se indican algunos inconvenientes que surgen al aplicar las *cartas de control* de manera tradicional, cuando las mediciones de un proceso están autocorrelacionadas. Posteriormente se incluyen preguntas de investigación, hipótesis, objetivos y justificación del problema.

2.1 Descripción del Problema.

Para utilizar una *carta de control clásica* se deben verificar dos supuestos: *normalidad e independencia* de las observaciones. Cuando los datos son *autocorrelacionados* o *serialmente correlacionados* no se cumple con la condición de independencia, es decir, existe memoria en los datos, por lo tanto las observaciones previas influyen a las observaciones futuras.

Ferrer (2004) menciona que cuando se presenta *autocorrelación positiva*, frecuentemente las cartas clásicas como *medias, rangos y varianzas* presentan un *ARL en control* inestable. Es decir, el desempeño de las cartas no es eficiente y se origina información errónea acerca de las *falsas alarmas*. Esto repercute en la interpretación de la carta de control, ya que la información que la carta arroja no es confiable. *Montgomery (2009)* menciona que estos inconvenientes se deben a que la varianza real del proceso se subestima.

Cuando los datos exhiben *autocorrelación negativa*, el efecto opuesto ocurre, es decir, se tiene una sobreestimación de la varianza real del proceso. Los límites de control son exageradamente amplios y se reduce la sensibilidad de detectar cambios reales en el proceso. Por consecuencia, ahora el *ARL fuera de control* aumenta, es decir, se requiere una gran cantidad de muestras en promedio para detectar un cambio de tamaño δ .

Cuando se construyen cartas de control, en un proceso con observaciones autocorrelacionadas, la estimación sesgada de la variación es un problema serio, que ocasiona una malinterpretación acerca las señales fuera de control. Por esto, se requiere desarrollar estrategias apropiadas para tratar esa estructura en los datos. Una carta modificada, que no presente problemas con frecuentes falsas alarmas o escasa potencia, y que permita detectar causas especiales; es la herramienta adecuada para los procesos industriales modernos, cuyas observaciones presentan autocorrelación.

La carta modificada propuesta debe ser comparada con la carta clásica, esto en términos de su *ARL*, con el fin de mostrar que el desempeño de la carta modificada es más eficiente que el de la carta clásica; cuando los datos presentan una estructura de autocorrelación. El *ARL en control*, ARL_0 , es el número de muestras promedio hasta que se detecta una falsa alarma; si se fija el error α (probabilidad de falsa alarma), se obtiene un ARL_0 de 370 para la carta de medias y de 200 para las cartas de varianzas y rangos. Sin embargo, con el efecto del parámetro de autocorrelación, ρ , y de cambios en el parámetro a monitorear, denotados como δ y medidos en unidades de desviación estándar; el ARL_0 se vuelve inestable, alejándose por encima o por debajo del valor objetivo, conforme ρ y δ aumentan. La mejor carta es aquella que tenga un ARL_0 estable, es decir, el valor más cercano a los valores de 370 o 200, para la carta de medias, rangos y varianzas.

2.2 Preguntas de Investigación

El incumplimiento de la suposición de independencia en las cartas de control motivó las siguientes preguntas:

¿Cuáles son los efectos de ignorar la correlación serial en las cartas de control clásicas?

¿Cuál carta de dispersión, resulta ser la más afectada en términos del ARL_0 por la autocorrelación en las observaciones?

¿Qué modificaciones se deben aplicar a una carta clásica con el fin de adecuarla a datos autocorrelacionados?

Lo anterior definió el punto de partida para la realización del presente proyecto de investigación.

2.3 Hipótesis y Variables de Investigación

El desempeño de las cartas de control para datos serialmente correlacionados es más eficiente, en términos del ARL_0 , que el desempeño de las cartas clásicas. Considerando el ARL_δ de la carta de control propuesta para datos autocorrelacionados se tiene que:

-Hipótesis Nula: El ARL_δ es inestable, mucho menor o mayor que el valor fijo del ARL_0 ; por lo tanto la carta tiene una tasa alta de falsas alarmas o la información no es confiable.

-Hipótesis Alterna: El ARL_δ es estable, muy cercano al valor fijo del ARL_0 ; por lo tanto la carta tiene una tasa baja de falsas alarmas.

2.4 Objetivo General

Desarrollar una metodología de monitorización eficiente para los procesos industriales modernos, que considere una carta de control específica para datos serialmente correlacionados.

2.4.1 Objetivos Específicos

-Estudiar los efectos de ignorar la correlación serial, para un proceso $AR(1)$, en las cartas clásicas.

-Modificar los límites de control de la carta clásica para medias y de una carta clásica de dispersión.

-Comparar el desempeño de las *cartas de control clásicas* con las *cartas de control basadas en el supuesto de correlación serial*, mediante su ARL_0 .

2.5 Justificación

2.5.1 Justificación Teórica

Como se menciona en el manual *AIAG de CEP (2005)*, en la industria moderna, específicamente en los procesos de conformado de la industria automotriz, la suposición de independencia es difícil de cumplir debido a que un comportamiento dinámico se presenta en la mayoría de estos procesos, los cuales tienden a ser principalmente dependientes de la maquinaria y la materia prima y en menor medida de otros factores. Este fenómeno ocasiona la presencia de correlación entre observaciones.

El presente documento genera una reflexión acerca la teoría clásica de las cartas de control. Estudiar los efectos de ignorar la correlación serial en las cartas permite realizar aportaciones al manejo y uso eficiente de las mismas, con el fin de eliminar problemas como frecuentes falsas alarmas o escasa potencia. Cuando se utiliza una carta de control se debe ser consciente de la parte teórica de su construcción, de manera que arroje resultados prácticos, confiables y fáciles de interpretar.

2.5.2 Justificación Práctica

Este documento aborda un caso de estudio en la empresa Sanden Mexicana S.A. de C.V., donde se identifica la causa de la correlación serial y se proponen estrategias que contribuyen a una monitorización eficiente en este tipo de procesos.

Debido a la prioridad de conocer las causas que ocasionan la correlación serial, es importante generar estrategias para identificar los elementos inerciales del proceso. Estrategias como análisis de la materia prima y el estudio de los parámetros adecuados, entre otras, son de utilidad para que la industria conozca y describa ampliamente sus procesos.

2.6 Delimitaciones.

El proyecto está enfocado a desarrollar una metodología de monitorización eficiente para los procesos productivos que presentan correlación serial, específicamente para aquellos que se explican con un *modelo autorregresivo, AR(1)*, aplicada en un proceso de conformado de metales.

CAPÍTULO 3. MARCO DE REFERENCIA

3.1 Marco Teórico

Una carta de control clásica debe cumplir dos supuestos para su implementación: normalidad e independencia de las observaciones. De acuerdo con *Harris y Ross (1991)*, la mayoría de las técnicas del *CEP* que son usadas para detectar cambios en la media o dispersión, asumen que las observaciones son independientes.

Alwan y Roberts (1995) estudiaron 235 aplicaciones de cartas de control y detectaron una mala colocación de los límites. La explicación que le encontraron a esta situación fue la violación del supuesto de independencia.

La condición de independencia comúnmente no se cumple en la práctica. *Ferrer (2004)*, comenta que la presencia de elementos inerciales, tales 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 es lo que ocasiona la presencia de *correlación serial o autocorrelación* en las mediciones de las características críticas del producto.

Wardell et al. (1992) y *AIAG Statistical Process Control (2005)* indican que muchos procesos industriales tales como las operaciones de maquinado y de forja presentan correlación serial. Bajo estas condiciones el *CEP* tradicional puede ser inapropiado para monitorear la calidad del proceso y se requieren métodos más adecuados. *Messaoud et al. (2005)* y *Elevli et al. (2009)* entre otros, utilizan dos cartas diferentes para tratar la autocorrelación en sus procesos, el primero en un proceso de manufactura de un producto y el segundo en un proceso de minería. Además los procesos autocorrelacionados son frecuentes en la industria química, automotriz y de servicios.

Nembhard et al. (2000) señalan que la autocorrelación, especialmente en muestras de unidades producidas muy cercanas en el tiempo, puede ser el resultado de muchos factores, tales como el cambio de turno, la rotación de operadores, cambios mecánicos o técnicos y funciones dentro de la misma maquinaria. *Haridy y Wu (2009)*, demuestran que la autocorrelación deteriora la habilidad de las cartas clásicas para distinguir correctamente las causas asignables de las causas comunes.

Young y Winistorfer (2001) y *Noffsinger y Anderson (2002)*, afirman que el *CEP* ha evolucionado a través del tiempo, ya que anteriormente se graficaban los datos históricos de corridas de producción de días o semanas anteriores. Actualmente en algunas empresas, los operadores recolectan mediciones en tiempo real, lo que les proporciona un conocimiento inmediato de las variables críticas del proceso que afectan la producción. Sin embargo, un inconveniente de tomar datos en esta forma, es que los procesos continuos pueden producir observaciones autocorrelacionadas. Es decir, las muestras en tiempo real tienden a tomar valores similares (autocorrelación positiva), lo que ocasiona que el proceso se aleje de su media u objetivo.

Vermaat (2006) menciona que en la actualidad el sector manufacturero se asemeja en poco al entorno donde las cartas de control tradicionales fueron introducidas. Debido a los sensores automatizados, en la actualidad la tasa de muestreo es a menudo muy alta, provocando autocorrelación. Por otro lado, existe un acuerdo general en cuanto a que la definición de control estadístico implica que la media y la desviación estándar del proceso permanezcan constantes en el tiempo. Una serie de tiempo estacionaria tiene una distribución de probabilidad con los primeros dos momentos constantes. Por lo tanto, se podría considerar que un proceso estacionario autocorrelacionado está en control estadístico.

Cheng y Thaga (2005) señalan que en algunas ocasiones la dinámica del proceso puede inducir correlación en observaciones que están muy cercanamente espaciadas en el tiempo. Si el intervalo de muestreo usado para monitorear el proceso es lo bastante corto para la dinámica del proceso, entonces se produce correlación serial significativa.

Por su cuenta *Runger y Willemain (1996)*, afirman que el incremento de sistemas de adquisición de datos automatizados ha producido volúmenes sin precedentes de datos con alta autocorrelación. Este patrón es más pronunciado en la industria química y de procesos, particularmente en conjunción con el uso de sistemas de control digital.

Es evidente que en la mayoría de las ocasiones la *correlación serial* o *autocorrelación* es ignorada. *Wieringa (1999)*, indica que esto ocasiona una interpretación deficiente de la información presentada en la carta de control. Debido a la *correlación serial* la desviación estándar del proceso es sobreestimada o subestimada provocando límites de control más amplios o reducidos, respectivamente.

Apley (2002) indica que el mal comportamiento de las cartas de control también afecta la curva del ARL_0 , la cual no permanece estable, sino que aumenta o disminuye notablemente dependiendo del tipo de carta que se utilice y del valor del parámetro de autocorrelación. Por lo general, la *autocorrelación positiva* incrementa el número de falsas alarmas y la *autocorrelación negativa* por su parte disminuye la sensibilidad, provocando que las causas especiales de variación no se detecten.

Por su parte, *Mason y Young (2002)* y *Shumway y Stoffer (2006)* afirman que la autocorrelación o correlación serial de las series de tiempo puede ser representada a través de modelos autorregresivos (AR) y con promedios móviles autorregresivos (ARMA). Adicionalmente se utiliza el modelo ARIMA (promedios móviles autorregresivos integrados) para procesos no estacionarios.

Young y Winistorfer (2001) aseguran que la *carta clásica de Shewhart* y la *carta de Promedios Móviles Ponderados Exponencialmente (EWMA)* para datos independientes arrojan una gran cantidad de falsas alarmas cuando las observaciones presentan *correlación serial positiva*. Ellos proponen la *carta ajustada por Wheeler para datos individuales* que es adecuada para un *proceso estacionario* representado por un modelo autorregresivo de primer orden $AR(1)$. Adicionalmente recomiendan la *carta EWMA ajustada* para procesos *no estacionarios*.

Noffsinger y Anderson (2002), mencionan algunos procedimientos para tratar datos autocorrelacionados, entre éstos destacan las *cartas para residuales* que utilizan modelos que se ajustan a los datos con el fin de graficar los residuales en la carta. Las *cartas para residuales* son recomendables para tratar observaciones que provienen de un *proceso no estacionario*. Sin embargo, presentan la desventaja de que en la práctica, es complicado ajustar

modelos a las observaciones en el tiempo real del proceso. Además las *cartas para residuales* no tienen las mismas propiedades que las *cartas clásicas* y en algunas ocasiones decrece su capacidad para detectar causas especiales de variación. Así mismo, *Apley (2002)*, comenta que cuando se trabaja con *cartas de residuales*, es común que existan errores de estimación en los parámetros del modelo.

Vermaat (2006), afirma que es recomendable utilizar *cartas tradicionales modificadas* ya que se evita el entrenamiento y el temor a manipular nuevas herramientas. Además comenta que en cuanto a las *cartas para residuales*, es complicado interpretar los residuales ya que no están en la escala original de las observaciones. Debido a esto, es mejor trabajar con un proceso que tiene datos autocorrelacionados y no transformarlo, ya que la condición de *estacionariedad* es suficiente para decir que el proceso está en control. Por lo tanto, el objetivo de la carta tradicional modificada será monitorear el proceso para asegurarse de que permanezca estacionario, en caso contrario se presentará una alarma.

Como se mencionó en el párrafo anterior, uno de los enfoques utilizados actualmente es modificar la carta clásica para tratar adecuadamente los datos correlacionados. El primer artículo escrito con este enfoque fue *Vasilopoulos y Stamboulis (1978)*. Tiempo después, *Schmid (1995)* analizó la conducta de la longitud de corrida de la carta modificada de Shewhart. Además, él comparó este esquema con la carta clásica de Shewhart aplicada a los residuales. La carta *EWMA* de *Roberts (1959)* fue extendida para datos correlacionados en *Schmid (1997a)*. Cartas modificadas *CUSUM* fueron discutidas por *Yashcin (1993)* y *Schmid (1997b)*.

De acuerdo con *Chang y Wu (2009)*, cuando un proceso está en control es deseable que la carta tenga un *ARL en control* grande. De otra manera, se

prefiere un *ARL fuera de control* pequeño cuando esté presente una señal fuera de control. Estos autores comentan que la autocorrelación debe ser eliminada, pero si ésta es parte inherente del proceso, debe ser modelada y monitoreada apropiadamente.

Kramer y Schmid (1998) afirman que para un proceso autorregresivo *AR(1)* es necesario estimar los coeficientes α_1^* y σ (parámetro de correlación y desviación estándar respectivamente del proceso). Una pobre estimación de α_1^* no tiene efectos severos en la curva de *ARL*, debido a que el *ARL en control* es casi una función simétrica con respecto a α_1 . Sin embargo, para la *carta modificada de Shewhart*, es importante contar con una buena estimación de la desviación estándar del proceso.

Por su parte *Cheng y Thaga (2005)* recomiendan que cuando el nivel de autocorrelación sea moderado, se utilicen cartas directamente para las observaciones. Si el nivel de autocorrelación es alto, es conveniente utilizar algún procedimiento de transformación para crear residuales. Si ocurre un cambio en la media o desviación estándar del proceso, de la misma manera habrá un cambio en la media o desviación estándar de los residuales.

En muchas aplicaciones prácticas con datos provenientes de un proceso *AR(1)*, es mejor tener tamaños de muestras grandes que espaciar la cantidad de observaciones entre una muestra y otra, excepto para los casos donde se requiere detectar cambios grandes en la media del proceso. *Runger y Willemain (1996)* sugieren un método donde los promedios de las observaciones autocorrelacionadas son tratados como individuales en una carta clásica para observaciones individuales. En este sentido, con el fin de aminorar la autocorrelación, se compara el desempeño de la carta en base a combinaciones formadas por: diferentes tamaños de subgrupos y espacios de diferentes cantidades de observaciones entre cada subgrupo; hasta obtener

una transformación conveniente para los datos originales. Para su investigación, *Knoth y Schmid (2002)*, asumen que cuando se utilizan cartas para subgrupos, las muestras pueden ser independientes entre ellas, pero las variables dentro de los mismos pueden ser correlacionadas. La motivación para introducir este modelo es el hecho de que el período entre dos muestras es frecuentemente más grande que el tiempo necesario para tomar una muestra.

Por otro lado, *Claro et al. (2008)* proponen una estrategia de doble muestreo, con el fin de recuperar la pérdida de eficiencia en las cartas de control; que es ocasionada por la autocorrelación. El procedimiento de doble muestreo asume que una muestra maestra n está formada por dos submuestras consecutivas ($n_1 + n_2$), que pueden ser recolectadas sin ningún efecto del tiempo, por lo tanto, estas submuestras provienen de la misma distribución de probabilidad.

En la literatura también se han estudiado índices de capacidad adecuados a procesos con datos serialmente correlacionados en: *Kotz y Johnson (1993)*, *Ramos y Ho (2003)*, *Lovelace et al. (2009)* y *Sun et al. (2010)*, lo que ha mostrado que la autocorrelación puede tener un serio impacto en la estimación de los índices de capacidad.

En *Psarakis (2007)*, se presentan, estudian y evalúan las cartas de control que han sido diseñadas para tomar en cuenta la autocorrelación. El autor concluye que todas las cartas mencionadas tienen sus ventajas y desventajas, pero que es de suma importancia construir cartas simples y efectivas para su pronta aplicación en los procesos industriales.

3.2 Marco Conceptual

3.2.1 Variación en los Procesos

En los primeros dos recuadros de la Fig. 0.1 se indica cómo el pensamiento estadístico se convierte en un enfoque lógico a seguir en todas las actividades que tienen el fin de mejorar la calidad. Todo el trabajo que se realiza puede ser visto como un proceso. Un proceso se define como (ver *Wieringa (1999)*): “un conjunto de causas y condiciones que actúan de manera interrelacionada para transformar entradas en salidas”. Un análisis exhaustivo de la variación, en combinación con el conocimiento del proceso que se tenga, puede conducir a la reducción de la variación.

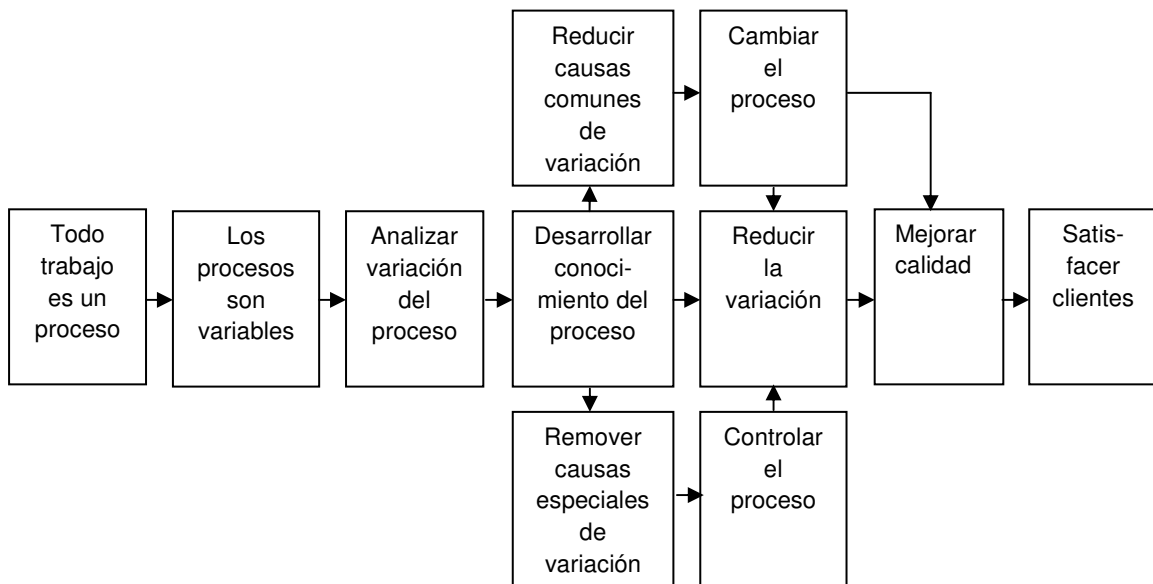


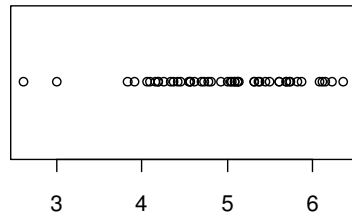
Fig. 0.1 Pensamiento Estadístico para la Mejora de la Calidad.

Un producto es de calidad si reúne las especificaciones del cliente. *Montgomery (2009)* afirma que calidad es inversamente proporcional a variabilidad. Esta definición implica que la mejora de la calidad se logra mediante la reducción de la variabilidad.

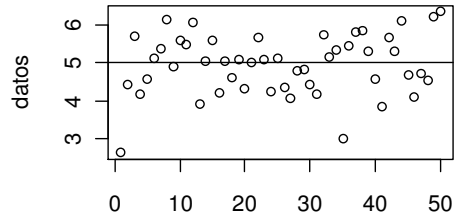
Wieringa (1999), reconoce dos tipos de variación. La variación debido a *causas comunes* es parte inherente del proceso, consecuencia de cómo ha sido diseñado; si sólo las causas comunes de variación están presentes, se dice que el proceso está estadísticamente *en control*. Lo anterior no significa que no exista variación o que la variación es pequeña, significa que las observaciones son predecibles dentro de los límites de control. Por otro lado, las *causas especiales o asignables* de variación no forman parte del proceso y ocurren solamente de forma esporádica o accidental. Este tipo de variación es ocasionada por un ajuste inapropiado de la maquinaria, errores del operador o materia prima defectuosa.

La reducción de la variación se lleva a cabo mediante dos formas. Una forma es mediante la remoción de causas especiales de variación, lo cual es responsabilidad de las personas que trabajan directamente en el proceso. Otra forma es reducir los efectos de las causas comunes de variación, lo cual requiere decisiones por parte de los administradores del proceso.

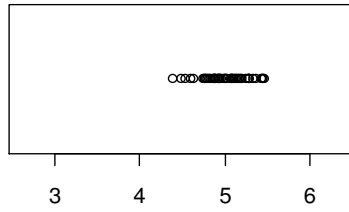
De acuerdo con Moore (2000) la descripción de una distribución de probabilidad con la que se modela un proceso de producción, exige una *medida de centralidad y una medida de dispersión*. Por lo que no basta con controlar la media u objetivo de un proceso de producción, también se requiere controlar su variabilidad. Es por lo anterior que para monitorear la calidad de un producto se controla la *media y la dispersión* de una característica de calidad de interés. La media es una medida de tendencia central o localización y la desviación estándar es un ejemplo de una medida de dispersión del proceso. A continuación, se representan datos en una gráfica de puntos y en una de dispersión. En la Fig. 0.2 se aprecia que los datos tienen desviación estándar diferente $\sigma = 0.7$ y $\sigma = 25$, pero la misma media $\mu = 5$. En la Fig. 0.3 se aprecia que los datos tienen medias diferentes $\mu = 5$ y $\mu = 5.25$, pero igual desviación estándar $\sigma = 0.25$.



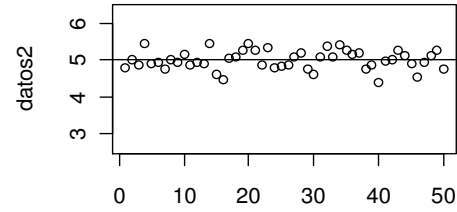
$\sigma = 0.7$



Número de observación
 $\sigma = 0.7$

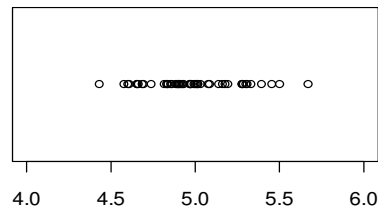


$\sigma = 0.25$

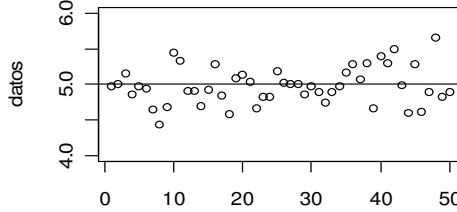


Número de observación
 $\sigma = 0.25$

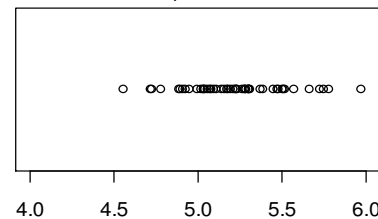
Fig. 0.2 Gráficas de Puntos y Dispersión con diferente Desviación Estándar.



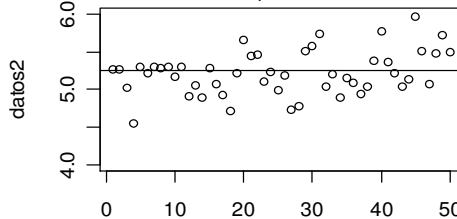
$\mu = 5$



Número de observación
 $\mu = 5$



$\mu = 5.25$



Número de observación
 $\mu = 5.25$

Fig. 0.3 Gráficas de Puntos y Dispersión con diferente Media.

3.2.2 Control Estadístico del Proceso: Cartas de control

Montgomery (2009) indica que la *carta de control* es una gráfica donde se representan una secuencia de estadísticos conformados por observaciones agrupadas o individuales. Tales observaciones se comparan con los límites de control, los cuales indican la magnitud de la variabilidad del proceso cuando sólo se presentan causas comunes de variación. Mientras todos los puntos estén dentro de los límites de control, se asume que el proceso está estadísticamente *en control*. En caso contrario, si un punto se encuentra fuera de los límites, (*señal fuera de control*), indica que hay más variación presente que la que se puede atribuir al efecto de las causas comunes. Sin embargo, debido a la condición aleatoria de las observaciones, existe una probabilidad pequeña de que una señal fuera de control aparezca mientras el proceso está *en control*. Tal señal se conoce como *falsa alarma*.

Di Bucchianico (2008) señala que es importante notar que existen otras anomalías en el proceso que puede provocar señales de alarma. Con el fin de incrementar el desempeño de detección de la carta, se pueden implementar reglas de corrida adicionales. Ejemplos de estas reglas incluyen patrones como ochos o más puntos sucesivos arriba o por debajo de la línea central, u ocho o más puntos incrementando o decreciendo en forma de fila, adhesión de los puntos a los límites de control y periodicidad.

Requerimientos de una Carta de Control

Para construir una carta de control se requiere especificar los *límites de control*, definir el *tamaño de la muestra* y la *frecuencia de muestreo*. Comúnmente los límites se ubican a tres desviaciones estándar de la línea

central. Estos límites son conocidos como límites de control 3-sigma. Considerando que los puntos fuera de control son la señal “fuera de control”, *Montgomery y Runger (2007)* comentan con respecto a los límites de control, cuando un punto fuera de los límites es la señal de alarma, se debe tener en consideración los siguientes puntos:

- Al alejar los límites de control de la línea central, se disminuye el *riesgo de un error tipo I*, es decir indicar una condición fuera de control cuando no está presente ninguna causa asignable.

-Si se acercan los límites de control a la línea central, aumenta el *riesgo del error tipo II*, es decir indicar que el proceso está en control, cuando existe una causa asignable.

En cuanto al *tamaño de muestra* se debe considerar la magnitud del corrimiento en la desviación estándar que se desea detectar. Para la frecuencia de muestreo por lo general se consideran muestras pequeñas a intervalos cortos o muestras más grandes a intervalos más largos.

Subgrupos Racionales

Para recolectar datos muestrales se utiliza el concepto de *subgrupos racionales*. *Juran et al. (1990)*, afirman que los subgrupos deben elegirse de forma que tengan mayor probabilidad de que las mediciones realizadas en cada subgrupo sean semejantes y de que los subgrupos se diferencien entre sí. *Montgomery y Runger (2007)* presentan tres enfoques para la formación de un subgrupo racional. El primero consiste en que las mediciones de cada subgrupo racional sean consecutivas o lo más cercanas posibles en el orden de producción. De manera que disminuyendo la variación dentro del subgrupo la

carta posee límites de control lo suficientemente sensibles como para detectar cambios en el proceso.

El segundo enfoque consiste en que las mediciones para cada subgrupo se tomen al azar. *Montgomery y Runger (2007)* afirman que este método suele emplearse para tomar decisiones acerca de la aceptación de todas las unidades que se han producido desde la última muestra. Sin embargo, cuando se toman mediciones al azar, la variación dentro de la muestra crece y los límites de control tienden a ser más grandes por lo que se hace difícil detectar corrimientos en la media.

Un tercer enfoque para la formación de subgrupos racionales es recolectar información para diferentes máquinas, materia prima, estaciones de trabajo, operadores, etc.. Aplicando una carta para diferentes elementos hace más fácil detectar de dónde proviene la causa especial de variación.

Construcción de una Carta de Control

En la *fase I* de la construcción de una carta de control se cuenta con observaciones denotadas como X_{ij} , donde $i = 1, \dots, m$, y $j = 1, \dots, n$. La cantidad m indica el número de la muestra y n el número de observaciones dentro de la misma. Se supone que todas las observaciones dentro y entre cada muestra son idénticas e independientemente distribuidas con una función de distribución F_i . La carta de control se basa en pruebas de hipótesis (3.1), con el fin de identificar si la función de distribución F_i cambia debido a causas especiales de variación:

$$\begin{aligned} H_0: F_1 = F_2 = \dots = F_m &\equiv F_0 \\ H_1: \text{Existe una } s, t \in \{1, 2, \dots, m\} &\text{ tal que } F_s \neq F_t \end{aligned} \quad (3.1)$$

La carta de control para un parámetro η_i de F_i se construye en base a que $T_i = T(X_{i1}, \dots, X_{in})$ es una estimación de η_i que se obtiene de una muestra y que se grafica en la carta de control.

Además se tiene que $M_m = M(X_{11}, \dots, X_{1n}, \dots, X_{m1}, \dots, X_{mn})$ es una estimación del parámetro de localización de la distribución de T_i bajo H_0 (tal como la media de las medias de las m muestras o media muestral global, que estima a la media poblacional) y $V_m = V(X_{11}, \dots, X_{1n}, \dots, X_{m1}, \dots, X_{mn})$ es un estimador del parámetro de dispersión de la distribución de T_i bajo H_0 .

Posteriormente, se comparan las observaciones de T_i contra los límites de control definidos de la forma $L = M_m \pm Z_{\alpha/2}V_m$. La constante $Z_{\alpha/2}$ comúnmente se reemplaza por el número tres. Si una observación de T_i se encuentra fuera de los límites de control se genera una alarma que se debe investigar. La fase I concluye hasta que no se encuentran falsas alarmas y se tiene una estimación de la distribución en control basadas en las m muestras denotada como F_0 . Para la fase II, en cada instante del tiempo una muestra X_{f1}, \dots, X_{fm} es recolectada por lo que se prueba la hipótesis (3.2):

$$H_0: F_f = F_0; \text{ contra la alternativa } H_1: F_f \neq F_0 \quad (3.2)$$

Estimadores para la Media

Wackerly et al. (2008) menciona que si el valor de un parámetro correspondiente a una variable de interés no se conoce, se usa un *estimador puntual*, mediante el cual se calcula el valor numérico o estimación puntual en base a las mediciones contenidas en una muestra.

Considerando una muestra de tamaño n de la población con observaciones X_1, X_2, \dots, X_n que son normales e independientes, se tiene que la *media muestral* (3.3) se define como la suma de todas las observaciones del subgrupo dividida por el tamaño n del mismo.

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n} \quad (3.3)$$

Así mismo, la *media muestral* tiene *media o valor esperado* y *varianza* que se muestran en (3.4) y (3.5), respectivamente. Por lo que la media muestral es un estimador puntual de la media poblacional μ .

$$\mu_{\bar{X}} = \frac{\mu + \mu + \dots + \mu}{n} = \frac{n\mu}{n}, \quad E(\bar{X}) = \mu \quad (3.4)$$

$$\sigma_{\bar{X}}^2 = \frac{\sigma^2 + \sigma^2 + \dots + \sigma^2}{n} = \frac{\sigma^2}{n} \quad (3.5)$$

Por ejemplo, para la *carta de medias* (\bar{X}), se utiliza μ como el límite central de la carta y $\mu \pm 3\sigma/\sqrt{n}$ como los límites superior e inferior. Si los valores correspondientes a los parámetros μ y σ^2 son desconocidos, se pueden estimar tomando m subgrupos racionales de tamaño n . Para estimar μ se calcula la *gran media o media global* (3.6) que es el promedio de las medias muestrales.

$$\bar{\bar{X}} = \frac{\bar{X}_1 + \bar{X}_2 + \dots + \bar{X}_m}{m} \quad (3.6)$$

Existen muchos estimadores, pero en cuanto a decidir cuál de todos es el mejor estimador puntual, *Wackerly et al. (2008)* afirma que se requiere evaluar si es *insesgado, consistente, eficiente y suficiente*. Para estimar μ se puede usar X_i ó \bar{X} . El estadístico $\bar{\bar{X}}$ es mejor estimador del parámetro

poblacional μ que X_i . En los siguientes párrafos se comparan los dos estimadores en base a sus propiedades.

Primero, un estimador es *insesgado* porque está próximo al valor del parámetro desconocido (Montgomery y Runger, 2007). En el caso de X_i y \bar{X} ambos son próximos a μ y son insesgados. Dado que no existe un estimador insesgado único se requiere evaluar las propiedades restantes para determinar cuál de los dos es el mejor estimador.

Segundo, según Wackerly et al. (2008) un estimador puntual es *consistente* si su valor tiende a estar más cerca del parámetro poblacional a medida que el tamaño de la muestra aumenta. Esta propiedad se cumple para el estimador \bar{X} .

Tercero, se sabe que la gran media, \bar{X} , y una observación de la muestra, X_i , son estimadores insesgados de μ . Para determinar cuál es más *eficiente* se calcula la eficiencia relativa (3.7). Después de calcularla, se observa que el estimador \bar{X} es más eficiente que el estimador X_i , dado que $\frac{1}{n} < n$.

$$\frac{\sigma_{\bar{X}}^2}{\sigma_{X_i}^2} < \frac{\sigma_{X_i}^2}{\sigma_{\bar{X}}^2} = \frac{\sigma^2/n}{\sigma^2} < \frac{\sigma^2}{\sigma^2/n} = \frac{1}{n} < n \quad (3.7)$$

Por último se asume que el estimador \bar{X} es *suficiente* dado que agota toda la información que contiene la muestra. En cambio X_i no es suficiente porque solo contiene una observación de toda la muestra.

Estimadores para la Desviación Estándar

De acuerdo con *Bartés et al. (2000)*, la estimación de σ se puede realizar a través de los *métodos de la desviación estándar o del rango* de cada muestra. Debido a su simplicidad, el método más utilizado en la práctica es el del rango.

Método de la Desviación Estándar

Se tienen m subgrupos racionales de tamaño n donde X_{ij} tiene $i = 1, \dots, m$, y $j = 1, \dots, n$, si las observaciones son independientes y distribuidas normalmente con media μ y varianza σ^2 , el estadístico $(n-1)s_n^2/\sigma^2$ se distribuye como $\chi_{(n-1)}^2$. Para obtener el valor esperado de s_n^2 (3.8) se tiene que el valor esperado de una ji-cuadrada son sus grados de libertad, de esta manera se concluye que $E(S_n^2) = \sigma^2$.

$$\begin{aligned} E[(n-1)s_n^2/\sigma^2] &= E[\chi_{(n-1)}^2] \\ (n-1)/\sigma^2 E(S_n^2) &= n-1 \end{aligned} \quad (3.8)$$

Por lo tanto, la *varianza muestral* es un estimador insesgado de la *varianza poblacional*. Sin embargo, la *desviación estándar muestral* S_n es un estimador sesgado de la *desviación estándar poblacional*. Para obtener un estimador insesgado de la desviación estándar poblacional se requiere un factor que corrija el sesgo. Si $t = \chi_{(n-1)}^2$, entonces \sqrt{t} es igual a la raíz cuadrada de la distribución ji-cuadrada con $(n-1)$ grados de libertad. De acuerdo con (3.8), el valor esperado de la *desviación estándar muestral*, $E[S_n]$, es:

$$E[S_n] = E[\sqrt{t}] * \sigma / \sqrt{n-1} \quad (3.9)$$

Según Kotz (1993) la función de densidad de la distribución ji-cuadrada, χ^2 , con parámetros x y k es la que se muestra en (3.10). El parámetro x equivale a t y k son $n - 1$ grados de libertad. La expresión $\Gamma(\alpha)$ es la función gamma definida en (3.11).

$$f(x; k) = \begin{cases} \frac{1}{2^{k/2} \Gamma(k/2)} x^{k/2-1} e^{-x/2} & \text{para } x \geq 0 \\ 0 & \text{para } x < 0 \end{cases} \quad (3.10)$$

$$\Gamma(\alpha) = \int_0^{\infty} y^{\alpha-1} e^{-y} dy \quad (3.11)$$

Utilizando (3.10), en (3.12) se inicia el procedimiento para encontrar el valor esperado correspondiente a la raíz cuadrada de la distribución ji-cuadrada, $E[\sqrt{t}]$.

$$\begin{aligned} E[\sqrt{t}] &= \int_0^{\infty} \frac{\sqrt{t}}{2^{n-1/2} \Gamma(n-1/2)} t^{(n-1/2)-1} * e^{(-t/2)} dt \\ &= \int_0^{\infty} \frac{1}{2^{n-1/2} \Gamma(n-1/2)} t^{(n/2)-1} * e^{(-t/2)} dt \end{aligned} \quad (3.12)$$

Para continuar en (3.13) se requiere un cambio de variable. De manera que $z = t/2$. El jacobiano dt/dz es igual a dos.

$$\begin{aligned} E[\sqrt{t}] &= \frac{1 * 2}{2^{n-1/2} \Gamma(n-1/2)} \int_0^{\infty} 2z^{(n/2)-1} * e^{(-z)} dz \\ &= \frac{1 * 2 * 2^{(n/2)}}{2 * 2^{n-1/2} \Gamma(n-1/2)} \int_0^{\infty} z^{(n/2)-1} * e^{(-z)} dz \end{aligned} \quad (3.13)$$

De acuerdo con (3.11), la expresión subrayada en (3.13) es una distribución gamma con parámetros $y = z$ y $\alpha = n/2$, por lo que se escribe de la forma $\Gamma(\alpha) = \Gamma(n/2)$. Finalmente como se muestra en (3.14), el procedimiento para obtener $E[\sqrt{t}]$, se ha simplificado.

$$\begin{aligned} E[\sqrt{t}] &= \frac{2^{(n/2)} \Gamma(n/2)}{2^{n-1/2} \Gamma(n-1/2)} \\ &= \frac{\sqrt{2} \Gamma(n/2)}{\Gamma(n-1/2)} \end{aligned} \quad (3.14)$$

Según (3.9), para obtener el $E[S_n]$ se requiere $E[\sqrt{t}]$, el cual se determinó en (3.14), la desviación estándar del proceso, σ , y el tamaño de muestra, n .

$$\begin{aligned} E[S_n] &= \frac{\sqrt{2} \Gamma(n/2)}{\Gamma(n-1/2)} * \frac{\sigma}{\sqrt{n-1}} \\ &= \frac{\sqrt{2} \Gamma(n/2)}{\Gamma(n-1/2)\sqrt{n-1}} * \sigma \end{aligned} \quad (3.15)$$

La expresión subrayada en (3.15) es la corrección para el sesgo, $C_4(n)$, la cual está en función del tamaño de muestra n . De esta manera, se obtiene que $E[S_n] = C_4(n) * \sigma$. Un *estimador insesgado* de σ se muestra en (3.16a). La varianza de S_n , $var(S_n)$, en (3.16b), es ahora fácil de calcular.

$$\hat{\sigma} = S_n / C_4(n) \quad (3.16a)$$

$$var(S_n) = E(S_n^2) - E^2(S_n) = \sigma^2 - C_4^2(n) \sigma^2 \quad (3.16b)$$

Simplificando, la varianza de S_n , es: $var(S_n) = (1 - C_4^2(n)) \sigma^2$. El *error estándar* o *desviación estándar* de S_n , es: $\sigma_{S_n} = \sigma \sqrt{1 - C_4^2(n)}$. En la Fig. 0.4 se

observa el comportamiento de $C_4(n)$ para diferentes tamaños de muestra. Mientras n sea una muestra pequeña $C_4(n)$ está cerca de uno, pero a medida que n aumenta de tamaño; $C_4(n)$ tiende a ser uno.

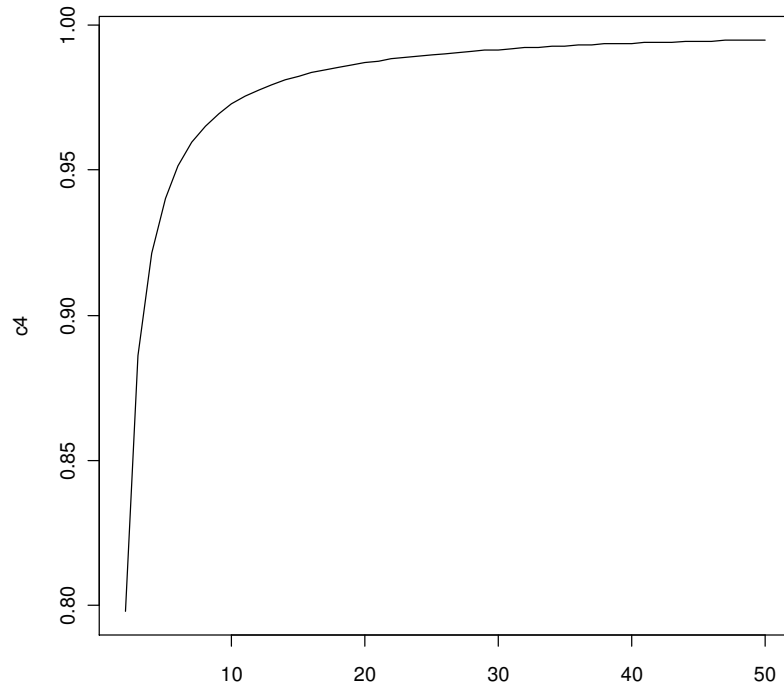


Fig. 0.4 Comportamiento de $c_4(n)$ para diferentes Tamaños de Muestra.

Método del Rango

Para calcular el *valor esperado del rango*, $E(R)$, en (3.17) se resta el valor esperado de la última observación (la máxima) y el valor esperado de la primera observación (la mínima).

$$E(R) = E[Y_n] - E[Y_1] \quad (3.17)$$

Por definición, el valor esperado de una variable aleatoria continua Y es: $E(Y) = \int_{-\infty}^{\infty} yf(y)dy$. (Wackerly et al, 2008). Por lo tanto, para calcular $E(R)$ se requiere la función de densidad de la n -ésima observación Y_n y de la primera Y_1 , mostradas en (3.18) y (3.19) respectivamente.

$$f_{Y_n}(Y) = n[F_x(Y)]^{n-1} * f_x(Y) \quad (3.18)$$

$$f_{Y_1}(Y) = n[1 - F_x(Y)]^{n-1} * f_x(Y) \quad (3.19)$$

Si requiere revisar el procedimiento realizado para obtener las funciones de densidad de (3.18) y (3.19), consulte el Apéndice al final del documento. De acuerdo con (3.17) el valor esperado del rango, $E(R)$, se obtiene de la forma:

$$\begin{aligned} E(R) &= E[Y_n] - E[Y_1] \\ &= \int_{-\infty}^{\infty} Y_n * nF_x^{n-1}(Y) * f_x(Y)dY - \int_{-\infty}^{\infty} Y_1 * n[1 - F_x(Y)]^{n-1} * f_x(Y)dY \\ &= \int_{-\infty}^{\infty} Y \frac{\partial}{\partial Y} [F_x^n(Y)dY] - \int_{-\infty}^{\infty} Y \frac{\partial}{\partial Y} [1 - [1 - F_x(Y)]^n] dY \\ &= \int_{-\infty}^{\infty} Y \frac{\partial}{\partial Y} [F_x^n(Y) - 1 + [1 - F_x(Y)]^n]dY \end{aligned} \quad (3.20)$$

En (3.21) se integra por partes la expresión obtenida de (3.20), de manera que se tiene que: $u = y, du = dy; v = [F_x^n(Y) - 1 + [1 - F_x(Y)]^n], dv = \frac{\partial}{\partial Y} [F_x^n(Y) - 1 + [1 - F_x(Y)]^n]$. Después de elegir adecuadamente los valores de u y dv , se simplifica la resolución de la integral:

$$\begin{aligned} E(R) &= Y[F_x^n(Y) - 1 + [1 - F_x(Y)]^n]_{-\infty}^{\infty} - \int_{-\infty}^{\infty} [F_x^n(Y) - 1 + [1 - F_x(Y)]^n]dY \\ &= \int_{-\infty}^{\infty} [1 - F_x^n(Y) - [1 - F_x(Y)]^n]dY \end{aligned} \quad (3.21)$$

De acuerdo con *Kiani y Paranetos (2008)*, para el caso de observaciones que provienen de una distribución normal, con media μ y varianza σ^2 conocidas, $F(x)$ es la función de distribución normal:

$$F(x) = \int_{-\infty}^x \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2} \frac{(y-\mu)^2}{\sigma^2}\right\} dy \quad (3.22)$$

Se utiliza una transformación de variable, de manera que $z = (y - \mu)/\sigma$, y el jacobiano dy/dz es igual a σ . Estandarizar la variable, como en (3.23) equivale a una transformación lineal, este procedimiento resulta útil para eliminar la dependencia de la variable respecto a las unidades de medida empleadas.

$$F(x) = \int_{-\infty}^{(y-\mu)/\sigma} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2} z^2\right\} dz = \Phi(z) = \Phi\left(\frac{y-\mu}{\sigma}\right) \quad (3.23)$$

De acuerdo a (3.21) y a la estandarización realizada en (3.23) se obtiene el *valor esperado del rango*, $E(R)$, mostrado en (3.24a). La cantidad $d_2(n)$ es el factor que corrige el sesgo y está en función del tamaño de muestra, n . De esta manera, se afirma que el estimador insesgado para la desviación estándar, se muestra en la ecuación (3.24b).

$$E(R) = \sigma \underbrace{\int_{-\infty}^{\infty} [1 - \Phi^n(z) - [1 - \Phi^n(z)]^n] dY}_{d_2(n)} \quad (3.24a)$$

$$\hat{\sigma} = \bar{R}/d_2(n) \quad (3.24b)$$

La cantidad $d_2(n)$ se calcula numéricamente para diferentes tamaños de n en el capítulo 5. En la Fig. 0.5 se grafica el comportamiento de $d_2(n)$ para diferentes tamaño de muestra n . A medida que n crece, $d_2(n)$ también se

incrementa. A diferencia de la constante $c_4(n)$ utilizada para el método de la desviación estándar, $d_2(n)$ tiene valores mayores a uno.

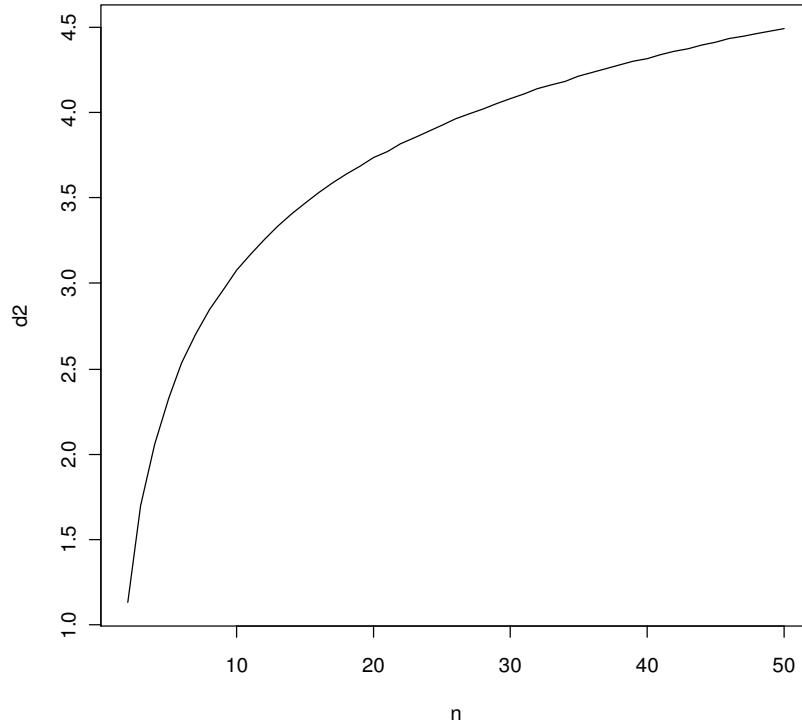


Fig. 0.5 Comportamiento de $d_2(n)$ para diferentes Tamaños de Muestra.

Se observa que el *valor esperado del rango muestral*, $E(R_i)$, es proporcional a la desviación estándar del proceso. Para obtener *la varianza del rango muestral* se utiliza la ecuación (3.25), donde $E(R_i)$ se obtuvo anteriormente y es igual a $d_2(n) * \sigma$, de acuerdo a la ecuación (3.24). Ahora se requiere obtener $E(R^2)$, de acuerdo a *Di Bucchianico (2008)*, $E(R^2)$ se calcula numéricamente mediante la doble integral mostrada en (3.26).

$$\text{var}(R_i) = E(R_i^2) - E^2(R_i) \quad (3.25)$$

$$E(R_i^2) = \int_0^{\infty} \int_{-\infty}^{\infty} r^2 * n(n-1) \{F(x+r) - F(x)\}^{n-2} f(x) f(x+r) dx dr \quad (3.26)$$

La doble integral de la ecuación (3.26) es igual a la siguiente expresión:

$$E(R_i^2) = [d_3^2(n) + d_2^2(n)] * \sigma^2 \quad (3.27)$$

Con el fin de obtener la *varianza del rango muestral* se utiliza la ecuación (3.25), donde de acuerdo a (3.24) y (3.27) la *varianza y desviación estándar del rango muestral* son:

$$\begin{aligned} var(R) &= [d_3^2(n) + d_2^2(n)] * \sigma^2 - d_2^2(n) * \sigma^2 \\ sd(R) &= d_3 * \sigma \end{aligned} \quad (3.28)$$

La cantidad $d_3(n)$ se calcula numéricamente para diferentes tamaños de muestra, n , en el capítulo 5 de este documento.

3.2.3 Cartas de Control Clásicas

Walter Shewhart propuso las *cartas de control clásicas* en 1931. Las cartas clásicas se clasifican en dos tipos. Las primeras se conocen como *cartas para variables* y son aquellas donde la característica de calidad es expresada como un número o en una escala continua de medición, por ejemplo diámetros, espesores y alturas. En cambio, las *cartas para atributos* son para aquellos productos que son clasificados como conformes o no conformes o para el conteo de defectos en una unidad de producción. En este documento se trata únicamente lo referente a las cartas para variables.

Carta para Medias

En una carta para medias, se grafica el promedio o media aritmética de cada muestra que se recolecta. Si se conocen los valores numéricos de los parámetros μ y σ , los límites superior e inferior de la carta son $\mu \pm 3\sigma/\sqrt{n}$ y el límite central es μ . Por otro lado, la estimación de la media poblacional μ está dada por la gran media, $\bar{\bar{X}}$:

$$\bar{\bar{X}} = \frac{1}{m} \sum_{i=1}^m \bar{X}_i \quad (3.29)$$

La gran media está definida como la suma de los promedios de cada muestra, dividida por el total de muestras, m , y se utiliza como el límite central de la carta de control para medias cuando el valor de μ no se conoce. Si se utiliza el método de rangos, σ se obtiene mediante la ecuación (3.24b). Por lo tanto, cuando se desconocen los valores numéricos de los parámetros, los límites de la carta de control para medias se representan como en (3.30).

$$L_{SC,IC} = \bar{\bar{X}} \pm \frac{3}{d_2\sqrt{n}} \bar{R}; \quad L_C = \bar{\bar{X}} \quad (3.30)$$

Si se utiliza el método de desviación estándar, mostrado en la ecuación (3.16a), para estimar σ , los límites de la carta de control para medias se obtienen mediante la ecuación (3.31a). En la práctica, los cálculos se facilitan utilizando las constantes que aparecen en la ecuación (3.31b), para diferentes tamaños de muestra.

$$L_{SC,IC} = \bar{\bar{X}} \pm \frac{3\bar{s}}{c_4\sqrt{n}}; \quad L_C = \bar{\bar{X}} \quad (3.31a)$$

$$A_2 = 3/(d_2\sqrt{n}); \quad A_3 = 3/(c_4\sqrt{n}) \quad (3.31b)$$

Carta para Rangos

En una muestra m de tamaño n , se tiene que la observación mayor es X_n y la menor se define como X_1 . El rango de cada muestra se calcula utilizando la ecuación (3.32). Para la carta de rangos se grafica el rango, R_i , de cada muestra que se recolecta.

$$R_i = X_n - X_1 \quad (3.32)$$

Si el valor de σ es conocido, la línea central de la carta es el *valor esperado del rango*, $E[R_i] = d_2 * \sigma$. Los límites de control superior e inferior de la carta para rangos son los que se muestran en (3.33a). Por el contrario, si se desconoce el valor de σ , se puede estimar mediante el método de rangos, $\hat{\sigma} = \bar{R}/d_2(n)$, y los límites superior e inferior se muestran en (3.33b).

$$L_{SC,IC} = d_2 \sigma \pm 3d_3 \sigma \quad (3.33a)$$

$$L_{SC,IC} = \bar{R} \pm 3d_3 \bar{R}/d_2 \quad (3.33b)$$

La estimación del rango poblacional está dada por el *rango promedio*, \bar{R} , definido como el promedio de los rangos de las m muestras:

$$\bar{R} = \frac{1}{m} \sum_{i=1}^m R_i \quad (3.34a)$$

Por lo tanto, el *rango promedio*, \bar{R} , se utiliza como la línea central de la carta de control para rangos cuando se desconoce el valor de σ . De acuerdo con (3.33a) y (3.33b), el calcular los límites de control de la carta requiere estimar la desviación estándar del rango $\hat{\sigma}_R$. Si los datos están distribuidos normalmente, $\hat{\sigma}_R$ se obtiene de la distribución del rango relativo $W = R/\sigma$.

Usando la distribución del rango relativo W , se tiene que $R = W\sigma$. De manera que la *desviación estándar del rango*, R_i , se obtiene mediante la ecuación (3.34b). En la práctica, la construcción de la carta de rangos se simplifica utilizando constantes. En este caso, $D_3 = 1 - 3\sigma_w/d_2$ y $D_4 = 1 + 3\sigma_w/d_2$.

$$\sigma_R = \sigma_W\sigma; \sigma_R = d_3\sigma \quad (3.34b)$$

Carta para Varianzas

Los límites de control probabilísticos para construir la carta de varianzas son los siguientes:

$$\begin{aligned} LIC &= \frac{\bar{s}^2}{n-1} \chi_{\alpha/2, n-1}^2 \\ LSC &= \frac{\bar{s}^2}{n-1} \chi_{1-(\alpha/2), n-1}^2 ; \quad LC = \bar{s}^2 \end{aligned} \quad (3.35)$$

donde $\chi_{\alpha/2, n-1}^2$ y $\chi_{1-(\alpha/2), n-1}^2$ corresponde a una distribución ji-cuadrada con $n-1$ grados de libertad. Si σ^2 es desconocido, \bar{s}^2 es el promedio de las varianzas obtenidas de muestras provenientes de una base de datos preliminar.

Carta para Desviaciones Estándar

En una carta de control clásica, se compara las observaciones contra los límites de control definidos de la forma $L = M \pm cV$. M es una estimación de la tendencia central del parámetro poblacional, V es una estimación de la dispersión del parámetro poblacional y c es una constante que comúnmente es

igual a tres. Para la carta de desviaciones estándar se grafica la desviación estándar, s_i , de las muestras recolectadas, (3.36).

$$s_i = \sqrt{\sum_{i=1}^m (x_i - \bar{x})^2 / n - 1} \quad (3.36)$$

Para el caso donde el valor de σ es conocido, la línea central de la carta es el *valor esperado de la desviación estándar*, $E[s_i] = c_4 * \sigma$. Con esta estimación se construyen los límites de control superior e inferior de la forma:

$$L_{SC,IC} = c_4 \sigma \pm 3\sigma \sqrt{1 - C_4^2(n)} \quad (3.37)$$

De otro modo si se desconoce el valor de σ , se puede estimar mediante el método de la desviación estándar, $\hat{\sigma} = \bar{S}/c_4$, y posteriormente se calculan los límites de control superior e inferior que se muestran en (3.38). En la práctica, los cálculos se simplifican utilizando las constantes: $B_3 = 1 - 3/c_4 \sqrt{1 - C_4^2(n)}$ y $B_4 = 1 + 3/c_4 \sqrt{1 - C_4^2(n)}$.

$$L_{SC,IC} = \bar{S} \pm 3 \bar{S}/c_4 \sqrt{1 - C_4^2(n)} \quad (3.38)$$

La estimación de la desviación estándar de la población está dada por la *desviación estándar promedio*, \bar{S} , (3.39), definida como el promedio de las desviaciones estándar de las m muestras.

$$\bar{S} = \frac{1}{m} \sum_{i=1}^m S_i \quad (3.39)$$

Por lo tanto, la desviación estándar promedio \bar{S} se utiliza como la línea central de la carta de control para desviaciones estándar cuando σ se desconoce. Calcular los límites de control de la carta requiere estimar la desviación estándar de S , definida como $\hat{\sigma}_{S_n}$. De acuerdo con lo expuesto en la sección *Método del Rango*, la *varianza de S* es: $var(S_n) = (1 - C_4^2(n)) \sigma^2$. Por lo tanto, la *desviación estándar de S* , es $\sigma_{S_n} = \sigma \sqrt{1 - C_4^2(n)}$.

3.2.4 Serie de Tiempo

Una *serie de tiempo* es una sucesión de variables aleatorias (proceso estocástico) indexadas por el tiempo t , $\{X_t : t \in T\}$. Suponiendo que se tiene una sucesión $x_0, x_1, \dots, x_n, \dots$ la observación x_i es recolectada en el instante i . En este documento el *conjunto índice* es discreto $T = \{0, 1, 2, \dots, n\}$. Las observaciones se recolectan de forma equiespaciada, es decir, en intervalos de tiempos iguales.

Utilizar el término “*serie de tiempo*” no es completamente adecuado para referirse al conjunto de observaciones tomadas de manera cronológica respecto al tiempo, ya que “*serie*” en matemáticas es utilizado para nombrar una suma infinita de observaciones de una variable. De acuerdo con *Guerrero (2003)* es más apropiado utilizar el término “*sucesión cronológica*”, sin embargo en este documento se escribe el término “*serie de tiempo*” por ser el más utilizado. Por medio de la representación gráfica de la *serie de tiempo* se puede detectar componentes de la misma, tales como *tendencias*, *variación estacional* o *cíclica* y *variación aleatoria*. La *tendencia* muestra la dirección de la variable a lo largo del tiempo, permite observar si ocurren cambios en la *media* de la serie. La *tendencia* puede ir continuamente en forma ascendente, descendente o permanecer fija, además puede ser lineal o no lineal. En la Fig. 0.6 se muestra una *serie de tiempo* con tendencia lineal ascendente. No se desea

este comportamiento en los procesos productivos, ya que en éstos se requiere mantener la variación oscilando alrededor de un objetivo, es decir, se prefiere que la tendencia permanezca fija.

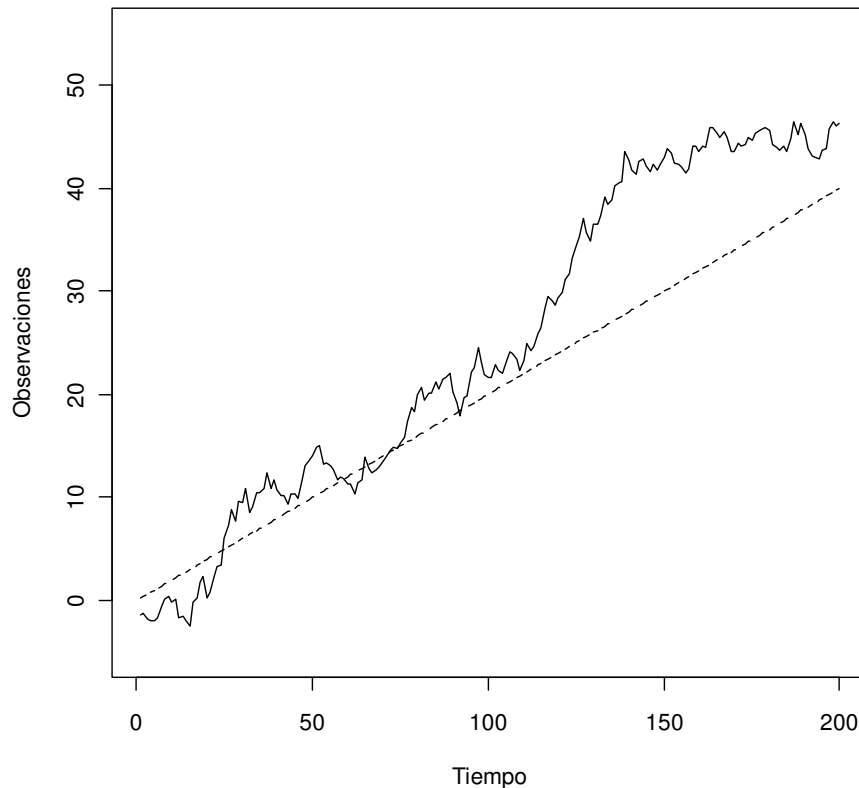


Fig. 0.6 Tendencia Ascendente en una Serie de Tiempo.

En una serie de tiempo se presenta *variación estacional* cuando el comportamiento de la variable se ve influenciado por las estaciones del año. Por ejemplo, el consumo de nieve se incrementa en verano, pero disminuye en invierno. Además, la serie de tiempo puede tener *variaciones cíclicas* que se caracterizan por oscilaciones periódicas de larga duración, (ver Fig. 0.7). Otro componente de la serie de tiempo es la *variación irregular* o ruido blanco, tal variación se presenta de manera aleatoria y por lo tanto no se puede predecir el efecto que tiene sobre la serie de tiempo.

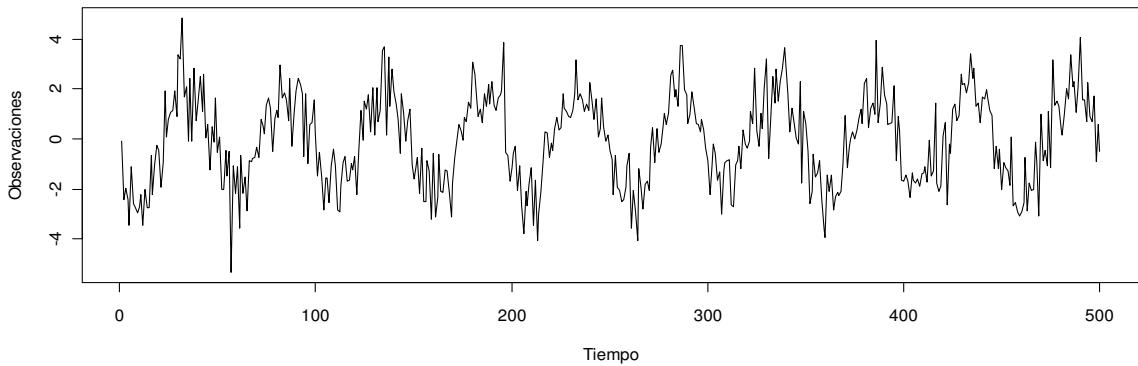


Fig. 0.7 Variación Cíclica en una Serie de Tiempo.

Ejemplos de Series de Tiempo.

Ruido Blanco Gaussiano.

Un *proceso estocástico* representado mediante variables aleatorias independientes e idénticamente distribuidas, con *media* 0 y *varianza* σ_w^2 , se conoce como *ruido blanco* y se denota como $W_t \sim iid(0, \sigma_w^2)$. Por otra parte, el *ruido blanco gaussiano*, se distingue del *ruido blanco* porque las W_t son variables aleatorias independientes con función de *distribución normal*. En la Fig. 0.8 se muestra un proceso de ruido blanco gaussiano con 100 observaciones, *media* 0 y *varianza* 1.

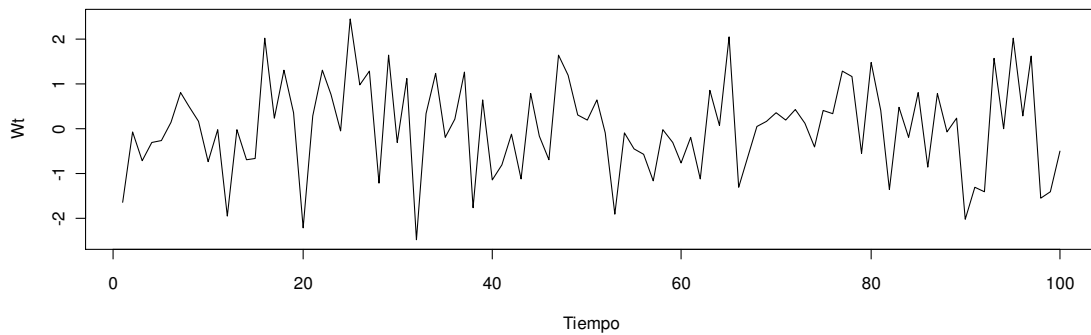


Fig. 0.8 Proceso de Ruido Blanco Gaussiano.

Promedios Móviles

Dada una serie X_t conformada por un conjunto de observaciones, la *serie de promedios móviles* v_t se genera mediante la ecuación:

$$v_t = \frac{1}{2p + 1} (x_{t-p} + x_{t-(p-1)} + \dots + x_{t-(p-n)} + x_t + x_{t+(p-n)} + x_{t+(p-(n-1))} + x_{t+(p-1)} + x_{t+p}) \quad (3.40)$$

donde el número de términos en la suma de la ecuación (3.40), $2p + 1$, corresponde a la cantidad de observaciones de la serie de promedios móviles. El término p , denota el orden de la serie, es decir, la memoria de la serie está limitada a p períodos. Se observa que se reemplaza la serie X_t por un *promedio móvil* v_t , tomando promedios de las observaciones hacia atrás y hacia adelante en base a la observación actual, x_t . En la

Fig. 0.9 se observa cómo la serie original mostrada en la Fig. 0.8, se suaviza aplicando un *promedio móvil de orden dos*, el cual se representa mediante la ecuación (3.41).

$$v_t = \frac{1}{2(2) + 1} (x_{t-2} + x_{t-1} + x_t + x_{t+1} + x_{t+2}) \quad (3.41)$$

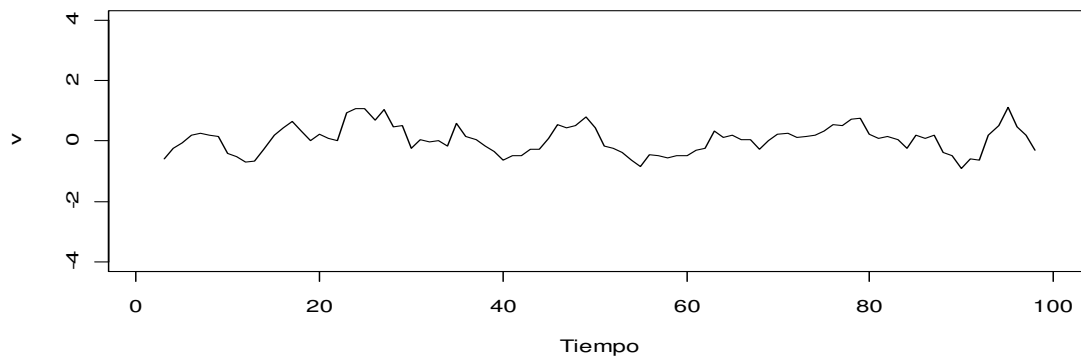


Fig. 0.9 Proceso de Promedios Móviles.

Autorregresiones.

Dada una serie X_t , conformada por un conjunto de observaciones, una *autorregresión* se representa mediante la ecuación:

$$x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + w_t \quad (3.42)$$

donde la observación x_{t-1} , es la observación anterior a la observación actual, x_t , y así hasta x_{t-p} , que es la observación a una distancia de p unidades en el tiempo de la observación actual, x_t . El número de retardos o regresores indican el orden p de la autorregresión. La ecuación (3.42) representa una regresión o predicción de la observación actual, x_t , como una función de sus observaciones anteriores en la serie.

La ecuación del *proceso autorregresivo* incorpora un término adicional, que es la parte no estructural del modelo y generalmente es un ruido blanco, w_t . En la

Fig. 0.10 se muestra un *proceso autorregresivo* de orden dos. Se tiene que $\mu = 0$, y se observa que por lo general, después de una observación por debajo de la media se presenta una observación por encima de la media y viceversa. Este patrón es característico de la *autocorrelación* o *correlación serial negativa*.

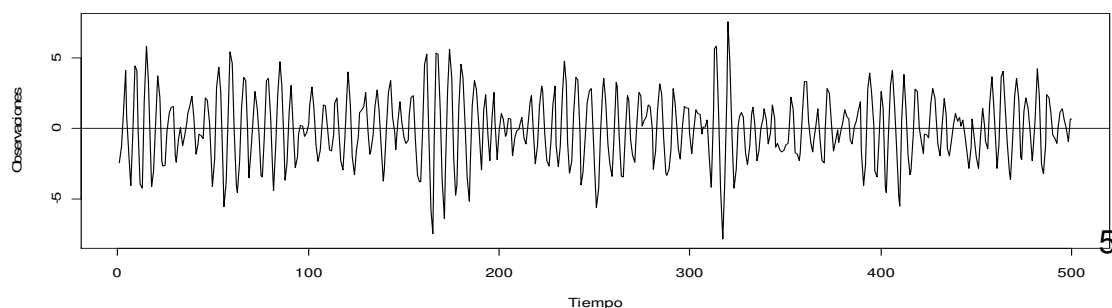


Fig. 0.10 Proceso Autorregresivo.

Caminata Aleatoria

Un modelo para analizar la tendencia de la serie de tiempo, es la *caminata aleatoria con rumbo*, representado de la forma:

$$x_t = \delta + x_{t-1} + w_t \quad (3.43)$$

para $t = 1, 2, \dots$, y donde w_t es el ruido blanco. La constante δ se conoce como el *rumbo* y cuando $\delta = 0$, el modelo se conoce como *caminata aleatoria simple*. Se utiliza el término *caminata aleatoria* ya que cuando $\delta = 0$, el valor de la serie de tiempo en el tiempo t es el valor de la serie en el tiempo $t - 1$ más un movimiento completamente aleatorio determinado por w_t .

Entonces, como se muestra en (3.44), otra manera de representar un *proceso de caminata aleatoria* es mediante una suma acumulada de las variables aleatorias de ruido blanco.

$$x_t = \delta t + \sum_{j=1}^t w_j \quad (3.44)$$

En la Fig. 0.11 se muestran 200 observaciones generadas de un *modelo de caminata aleatoria* con $\delta = 0$, $\delta = 0.3$ y con $\sigma_w = 1$. Se observa que cuando $\delta = 0.3$, la serie tiene rumbo; en cambio cuando $\delta = 0$, la serie es una *caminata aleatoria simple*; puesto que no tiene rumbo y permanece alrededor de la *media u objetivo del proceso*.

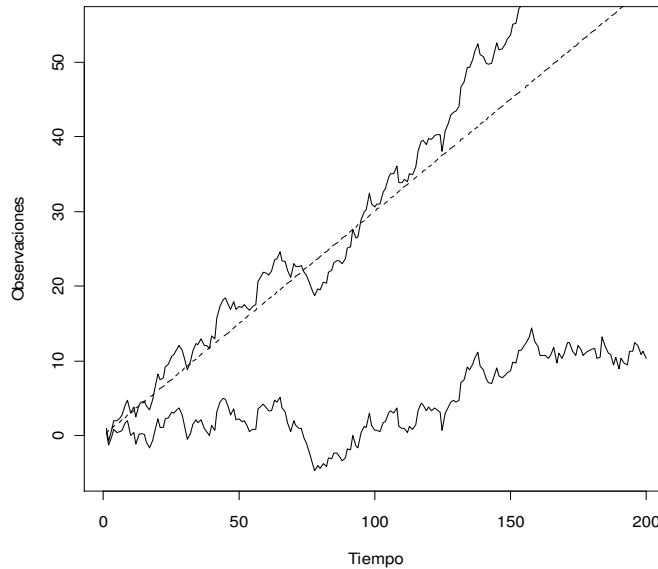


Fig. 0.11 Caminata Aleatoria.

3.2.5 Medidas de Dependencia: Autocorrelación

Función de Medias

La *función de medias* (3.45) de un proceso proporciona la media de x_t en un instante particular, t , del tiempo. Se dice que la media del proceso no depende del tiempo t , si la función de medias es una constante.

$$\mu_{xt} = E(x_t) = \int_{-\infty}^{\infty} x f_t(x) dx \quad (3.45)$$

Por ejemplo, se tienen dos procesos (3.46) y (3.47) y se pretende analizar su función de medias. Para el proceso (3.46) el *valor esperado de* x_t , $E(x_t)$, depende del tiempo, t . En cambio, para el proceso (3.47) el *valor esperado o media* es una constante y no depende del tiempo.

$$\begin{aligned}
x_t &= \beta_1 + \beta_2 t + w_t \\
E(x_t) &= E(\beta_1) + E(\beta_2 t) + E(w_t) \\
E(x_t) &= \beta_1 + \beta_2 t
\end{aligned} \tag{3.46}$$

$$\begin{aligned}
Y_t &= x_t - x_{t-1} \\
E(Y_t) &= E(x_t) - E(x_{t-1}) \\
E(Y_t) &= \beta_1 + \beta_2 t - (\beta_1 + \beta_2(t-1)) \\
E(Y_t) &= \beta_1 + \beta_2 t - \beta_1 - \beta_2(t-1) \\
E(Y_t) &= \beta_2 [t - (t-1)] \\
E(Y_t) &= \beta_2
\end{aligned} \tag{3.47}$$

En el caso de un *proceso de promedios móviles*, su *función de medias* es cero y no depende del tiempo como se observa en (3.48). En cambio de acuerdo con (3.49), examinando un *proceso de caminata aleatoria con rumbo*, se tiene que la *función de medias* depende del tiempo, si $\delta \neq 0$.

$$\mu_{vt} = E(v_t) = \frac{1}{3} [E(x_{t-1}) + E(x_t) + E(x_{t+1})] = 0 \tag{3.48}$$

$$\mu_{xt} = E(x_t) = \delta t + \sum_{j=1}^t w_j = \delta t \tag{3.49}$$

Función de Autocovarianza

La *autocovarianza* mide la dependencia lineal entre dos puntos de la misma serie, observados en diferentes instantes del tiempo, t , como se muestra en (3.50). De acuerdo con (3.51), para un *proceso de ruido blanco*, si $s = t$ la autocovarianza se reduce a la varianza.

$$\gamma_x(s, t) = E[(x_s - \mu_s)(x_t - \mu_t)] \tag{3.50}$$

$$\gamma_x(t, t) = E[(x_t - \mu_t)^2] \tag{3.51}$$

Como se muestra en (3.52), si $s \neq t$, la función de autocovarianza es igual a cero. Ahora considere un *promedio móvil de orden tres* aplicado a la *serie de ruido blanco*, w_t , como se muestra en (3.53).

$$\gamma_w(s, t) = E(w_s w_t) = \begin{cases} \sigma_w^2, & s = t \\ 0, & s \neq t \end{cases} \quad (3.52)$$

$$\begin{aligned} \gamma_v(s, t) &= E[(v_s - 0)(v_t - 0)] \\ &= \frac{1}{9} E[(w_{s-1} + w_s + w_{s+1})(w_{t-1} + w_t + w_{t+1})] \end{aligned} \quad (3.53)$$

Es conveniente obtener la *autocovarianza* en función de la separación, $s - t = h$, es decir, para $h = 0, \pm 1, \pm 2$. La *función de autocovarianza* para un *promedio móvil de orden tres* con $h = 1$ y $h = 2$ se muestra en (3.54) y (3.55), respectivamente.

$$\begin{aligned} \gamma_v(t, t) &= \frac{1}{9} E[(w_{t-1} + w_t + w_{t+1})(w_{t-1} + w_t + w_{t+1})] \\ &= \frac{1}{9} [E(w_{t-1} w_{t-1}) + E(w_t w_t) + E(w_{t+1} w_{t+1})] = \frac{3}{9} \end{aligned} \quad (3.54)$$

$$\begin{aligned} \gamma_v(t + 1, t) &= \frac{1}{9} E[(w_t + w_{t+1} + w_{t+2})(w_{t-1} + w_t + w_{t+1})] \\ &= \frac{1}{9} [E(w_t w_t) + E(w_{t+1} w_{t+1})] = \frac{2}{9} \end{aligned} \quad (3.55)$$

En la ecuación (3.56), se muestra que el *suavizamiento de la serie* ocasiona que la *función de autocovarianza* decrezca a medida de que la separación entre las dos observaciones se incremente y desaparezca completamente cuando las observaciones de la variable se separan a una distancia de tres o más puntos en el tiempo.

$$\gamma_v(s, t) = \begin{cases} 3/9, & s = t \\ 2/9, & |s - t| = 1 \\ 1/9, & |s - t| = 2 \\ 0, & |s - t| \geq 3. \end{cases} \quad (3.56)$$

Por último, considere un proceso de *caminata aleatoria*. Se observa que la *función de autocovarianza* de este proceso mostrada en (3.57), depende del tiempo en particular de las observaciones s y t y no del tiempo de la separación o retraso, h , a diferencia de los casos anteriormente explicados.

$$\gamma_x(s, t) = cov(x_s, x_t) = cov\left(\sum_{j=1}^s w_j, \sum_{k=1}^t w_k\right) = \min\{s, t\} \sigma_w^2 \quad (3.57)$$

Además, es importante considerar que la varianza de la *caminata aleatoria*, $var(x_t) = \gamma_x(t, t) = t \sigma_w^2$, está en función del tiempo, t . Eso se puede observar en la Fig. 0.11, donde la *varianza del proceso* aumenta a medida que el tiempo transcurre y la serie tiende a moverse lejos de su *función de medias* cuando $\delta = 0.3$.

Función de Autocorrelación

La *función de autocorrelación* mostrada en (3.58) mide la predictibilidad lineal de la serie en el tiempo t , es decir, la observación x_t , usando solamente las observación x_s . Si se predice x_t perfectamente de x_s a través de la relación lineal, $x_t = \beta_0 + \beta_1 x_s$, entonces la correlación será 1 cuando $\beta_1 > 0$, y -1 cuando $\beta_1 < 0$.

$$\rho(s, t) = \frac{\gamma(s, t)}{\sqrt{\gamma(s, s)\gamma(t, t)}} \quad (3.58)$$

Para simplificar la notación, es conveniente representar la *función de autocorrelación* en base a la diferencia, $h = |s - t|$, de manera que ahora se escribe:

$$\rho(h) = \frac{\gamma(t+h, t)}{\sqrt{\gamma(t+h, t+h)\gamma(t, t)}} = \frac{\gamma(h)}{\gamma(0)} \quad (3.59)$$

Para entender el concepto de autocorrelación, se requiere recordar la fórmula para estimar *el coeficiente de correlación simple* mostrada en (3.60), éste es un índice que mide la relación lineal entre dos variables aleatorias y y x .

$$r = \frac{\sum(y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum(y_i - \bar{y})^2 \sum(x_i - \bar{x})^2}} \quad (3.60)$$

De manera semejante, para la *función de autocorrelación* se tiene una serie de tiempo con n observaciones, es decir solamente una variable con la cual se pueden formar $(n - 1)$ parejas contiguas, por ejemplo: $(x_1, x_2), (x_2, x_3), \dots, (x_{n-1}, x_n)$, y si se estima el coeficiente de correlación, se obtiene el *coeficiente de autocorrelación de orden uno*, $\rho(1)$. Así mismo, se pueden formar parejas con observaciones separadas por una distancia de dos unidades de tiempo para estimar su correlación, por ejemplo: $(x_1, x_3), (x_2, x_4), \dots, (x_{n-2}, x_n)$, obteniéndose su *coeficiente de autocorrelación de orden dos*, $\rho(2)$.

La *función de autocorrelación*, (FAC), es el conjunto de los *coeficientes de autocorrelación*, $\rho(h)$, desde $h = 1$ hasta una cantidad que no exceda la mitad de las observaciones de la serie de tiempo. En la Fig. 0.12 se muestra un *correlograma* de un proceso que presenta *correlación serial o autocorrelación*, el eje horizontal representa el *retraso*, h , y el eje vertical representa el *coeficiente de autocorrelación*, $\rho(h)$, para cada retraso.

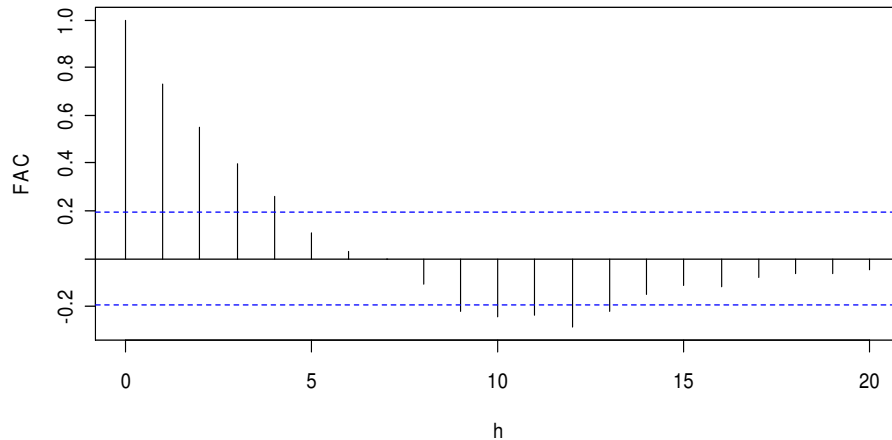


Fig. 0.12 Correlograma para una Serie de Tiempo.

De manera general, si w_t representa un ruido blanco, para un valor de n grande, la *FAC* muestral, $\hat{\rho}_x(h)$, para $h = 1, 2, \dots, H$, donde H es fijo pero arbitrario, está distribuida normalmente con media cero y desviación estándar $1/\sqrt{n}$. De esta forma, se obtiene un método aproximado para evaluar si cada $\hat{\rho}(h)$ es significativo, mediante el intervalo $\pm 2/\sqrt{n}$.

Si el valor numérico de la estimación correspondiente a los *coeficientes de autocorrelación*, $\rho(h)$, es superior o inferior al *intervalo de confianza*; tal valor es significativo, es decir los datos son autocorrelacionados. De igual forma, se puede construir un *diagrama de dispersión* para mostrar gráficamente cada una de las estimaciones de los *coeficientes de autocorrelación*, $\hat{\rho}(h)$.

3.2.6 Procesos Estacionarios

Proceso Estocástico Lineal

Un *proceso estocástico lineal*, x_t , se representa como en (3.61); en donde el parámetro μ determina la media del proceso y el modelo se basa en

que las observaciones sucesivas dependientes del tiempo están ponderadas por el parámetro ψ_j , y son generadas mediante un proceso de ruido blanco, w_t . De acuerdo con Guerrero (2003), para caracterizar por completo un *proceso estocástico* se requiere conocer la *función de densidad conjunta* de sus variables aleatorias. Sin embargo, en la práctica esto no se puede obtener para una serie de tiempo.

$$x_t = \mu + \sum_{-\infty}^{\infty} \psi_j w_{t-j} \quad (3.61)$$

Entonces, es conveniente obtener los momentos de primer y segundo orden, es decir, *medias, varianzas y covarianzas* del *proceso estocástico* en cuestión para conocer la forma de la distribución del mismo. Para esto se supone que la media de x_t es μ_t , es decir, $E(x_t) = \mu_t$. Entonces, se tiene que:

$$\mu_t = \mu + E(w_t - \psi_1 w_{t-1} - \psi_2 w_{t-2} - \dots) \quad (3.62)$$

Para que se pueda obtener la esperanza de la expresión (3.62), es necesario que el valor absoluto de la serie de ponderaciones converja, es decir que los coeficientes ψ_j satisfagan lo siguiente:

$$\psi_0 + \sum_{i=1}^{\infty} |\psi_j| < \infty \quad (3.63)$$

donde $\psi_0 = 1$, y de esta forma se tendría que $E(w_t) = 0$ para cualquier t . Si esto se cumple, la media del proceso, μ_t , no depende del tiempo. De manera que si en algún lapso de tiempo, el proceso se aleja de la media, en otro lapso tendrá que acercarse de nuevo a la misma. En la Fig. 0.13 se muestra un proceso con *media constante*, es decir su media no depende del tiempo.

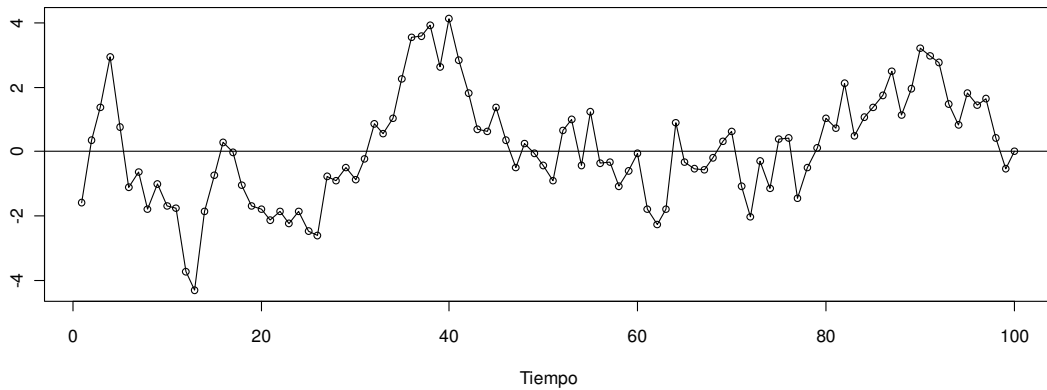


Fig. 0.13 Un Proceso con Media Constante.

De acuerdo con (3.61) la expresión $x_t - \mu$ es igual a la sumatoria de ruidos blancos, $\sum_{-\infty}^{\infty} \psi_j w_{t-j}$, ponderados por un parámetro ψ_j . De modo que para obtener la varianza del *proceso estocástico lineal* denotada como γ_0 , se utiliza la definición clásica de varianza, tal como se observa en (3.64).

$$\begin{aligned}
 \gamma_0 &= E[(x_t - \mu)^2] \\
 &= E[(w_t - \psi_1 w_{t-1} - \psi_2 w_{t-2} - \dots)^2] \\
 &= E(w_t^2 + \psi_1^2 w_{t-1}^2 + \psi_2^2 w_{t-2}^2 + \dots) - \\
 &\quad E(\text{productos cruzados de } \psi_i w_{t-i} \text{ y } \psi_j w_{t-j}) \\
 &= \sigma_w^2 \sum_{i=0}^{\infty} \psi_i^2 \tag{3.64}
 \end{aligned}$$

donde de acuerdo con (3.63) es necesario que la expresión $\sum_{i=0}^{\infty} \psi_i^2$ converja. El valor esperado de los productos cruzados en (3.64) desaparece ya que la variable aleatoria w_t es independiente con media cero.

Además cuando una serie de tiempo se considera como un proceso estocástico lineal se requiere estimar la covarianza entre x_t y x_{t+h} , la cual se escribe como $\gamma_h = cov(x_t, x_{t+h})$ y está dada por:

$$\begin{aligned}
\gamma_h &= E[(x_t - \mu)(x_{t+h} - \mu)] \\
&= E[(w_t - \psi_1 w_{t-1} - \psi_2 w_{t-2} - \dots)(w_{t+h} - \psi_1 w_{t+h-1} - \psi_2 w_{t+h-2} - \dots)] \\
&= E(-\psi_h w_t^2 + \psi_1 \psi_{h+1} w_{t-1}^2 + \psi_2 \psi_{h+2} w_{t-2}^2 + \dots) - \\
&\quad E(\text{productos cruzados de } \psi_i w_{t-i} \text{ y } \psi_j w_{t-j}) \\
&= \sigma_w^2 (-\psi_h + \psi_1 \psi_{h+1} + \psi_2 \psi_{h+2} + \dots) \\
&= \sigma_w^2 \left(\sum_{i=1}^{\infty} \psi_i \psi_{h+i} - \psi_h \right) \tag{3.65}
\end{aligned}$$

Procesos Débilmente Estacionarios

Los resultados obtenidos en (3.62) y (3.64) permiten definir una *serie débilmente estacionaria* o también conocida como *serie estacionaria de segundo orden*, la cual debe cumplir con los siguientes requisitos:

1. La función de medias, μ_t es constante y no depende del tiempo t .
2. Su variabilidad es finita y no depende del tiempo. Asimismo su covarianza no presenta dependencia del tiempo, pero si de la separación, h , que existen entre las variables.

Con lo anterior se puede afirmar que de manera general, la *serie de tiempo* muestra un comportamiento igual en el *primer y en el segundo momento* (medias, varianzas y covarianzas) sin importar el momento en el tiempo en que se observe. Dicho de otra forma, si se grafican observaciones contiguas de la serie, y luego otra cantidad de observaciones contiguas pero h períodos hacia adelante o atrás en el tiempo, ambas gráficas tendrán un aspecto similar.

Procesos Estacionarios en Sentido Estricto

En cambio, una *serie de tiempo estrictamente estacionaria* es aquella donde el comportamiento probabilístico de cada conjunto de valores, $\{x_{t1}, x_{t2}, \dots, x_{tk}\}$ es idéntico al comportamiento del conjunto desfasado en el tiempo, $\{x_{t1+h}, x_{t2+h}, \dots, x_{tk+h}\}$.

Shumway y Stoffer (2008) afirman que si la serie de tiempo es *estrictamente estacionaria*, entonces todas las funciones de distribución multivariadas para los subconjuntos de variables deben estar de acuerdo con sus contrapartes en el subconjunto desfasado para todos los valores del parámetro de desfase, h , de manera que:

$$f(x_t, x_{t+1}, \dots, x_{t+m}) = f(x_{t+h}, x_{t+h+1}, \dots, x_{t+h+m}) \forall t, m \text{ y } h. \quad (3.66)$$

Un *proceso estacionario en sentido estricto* también es estacionario en *sentido débil* o de *segundo orden*. No se puede afirmar lo contrario, a menos que todas las variables que componen la serie tengan una distribución conjunta *normal multivariada o gaussiana*.

Guerrero (2003), señala que es difícil, en la mayoría de las aplicaciones donde se cuenta con una base de datos preliminar, asegurar que un proceso es *estacionario en sentido estricto*. Es preferible utilizar el concepto de *estacionario de segundo orden* o en *sentido débil*, donde se imponen condiciones solamente para los primeros dos momentos de la serie de tiempo. Es decir, la función de medias y la función de autocovarianza son suficientes para caracterizar una serie estacionaria. En adelante, se utilizará el término *estacionario*, para referirse a una *serie estacionaria de segundo orden*.

Estimaciones para una Serie Estacionaria

Para caracterizar un *proceso estacionario* se estima la *función de medias muestral* (3.67), la *función de autocovarianza muestral* (3.68), y la *función de autocorrelación muestral* (3.69).

$$\hat{\mu} = \frac{1}{n} \sum_{j=1}^n x_j \quad (3.67)$$

$$\gamma(\hat{h}) = \frac{1}{n} \sum_{t=1}^{n-h} (x_{t+h} - \bar{x})(x_t - \bar{x}) \quad |h| = 0,1,2,3 \dots \quad (3.68)$$

$$\rho(\hat{h}) = \frac{\gamma(\hat{h})}{\gamma(0)} \quad (3.69)$$

donde es necesario que se satisfagan las condiciones dadas en la sección *Procesos débilmente estacionarios*. La *función de autocorrelación muestral* tiene una distribución muestral que permite saber si los datos son aleatorios o si los *coeficientes de correlación* son estadísticamente significativos para algún retraso, h , de la serie.

3.2.7 Modelos Autorregresivos AR(p)

Un proceso autorregresivo de orden p , conocido como $AR(p)$ se representa de la forma:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} \dots + \phi_p x_{t-p} + w_t \quad (3.70)$$

donde x_t es un proceso estacionario, $\phi_1, \phi_2, \dots, \phi_p$ son constantes diferentes de cero y w_t es un proceso de ruido blanco gaussiano con media

cero y varianza σ_w^2 . La variable aleatoria x_t depende de sus propios valores en los períodos de retraso $t - 1, t - 2, \dots, t - k$. La media del proceso x_t es cero, de lo contrario, se reemplaza x_t por $x_t - \mu$ en todo el modelo de la forma:

$$x_t - \mu = \phi_1(x_{t-1} - \mu) + \phi_2(x_{t-2} - \mu) \dots + \phi_p(x_{t-p} - \mu) + w_t \quad (3.71)$$

$$x_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} \dots + \phi_p x_{t-p} + w_t \quad (3.72)$$

en la ecuación (3.72) la constante c es igual a la cantidad $(1 - \phi_1 - \phi_2 - \dots - \phi_p)\mu$, siempre y cuando el proceso x_t , tenga una media constante dada por $E(x_t) = \mu$ para toda t . Adicionalmente, se tiene un modelo de regresión lineal, sin embargo, se utiliza el término *autorregresión* ya que se tiene una sola variable desfasada en el tiempo.

Otra forma de escribir un *proceso autorregresivo* es con el *operador de retraso* que se aplica sobre x_t y se muestra en (3.73). De manera que si tenemos $B(x_t)$ esto es igual a x_{t-1} , $B^2(x_t) = x_{t-2}$ y así hasta $B^P(x_t) = x_{t-P}$.

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)x_t = w_t$$

$$\phi(B)(x_t) = w_t \quad (3.73)$$

Proceso Autorregresivo de Orden 1, *AR(1)*

Un proceso *AR(1)* se representa mediante la ecuación (3.74). Suponiendo que la media es igual a cero, se tiene que $c = 0$. Iterando k veces hacia atrás se obtiene la expresión que se muestra en (3.75).

$$x_t = c + \phi x_{t-1} + w_t \quad (3.74)$$

$$x_t = \phi (\phi x_{t-2} + w_{t-1}) + w_t$$

$$\begin{aligned}
&= \phi^2 x_{t-2} + \phi w_{t-1} + w_t \\
&= \phi^2 (\phi x_{t-3} + w_{t-2}) + \phi w_{t-1} + w_t \\
&= \phi^3 x_{t-3} + \phi^2 w_{t-2} + \phi w_{t-1} + w_t \\
&= \phi^k x_{t-k} + \sum_{j=0}^{k-1} \phi^j w_{t-j} \tag{3.75}
\end{aligned}$$

Si se sigue iterando hacia atrás, cumpliendo con los requisitos de que el proceso x_t sea estacionario y que $|\phi| < 1$, el término $\phi^k x_{t-k}$ tiende a ser cero. Por lo tanto, el modelo $AR(1)$ puede representarse como una combinación lineal de ruidos blancos (3.76) ponderados por el parámetro ϕ . El proceso x_t además es estacionario ya que tiene función de medias constante (3.77), su función de autocovarianza depende de h en (3.78) y varianza finita en (3.79).

$$x_t = \sum_{j=0}^{\infty} \phi^j w_{t-j} \tag{3.76}$$

$$E(x_t) = \sum_{j=0}^{\infty} \phi^j E(w_{t-j}) = 0 \tag{3.77}$$

$$\begin{aligned}
\gamma(h) &= cov(x_{t+h}, x_t) = E \left[\left(\sum_{j=0}^{\infty} \phi^j w_{t+h-j} \right) \left(\sum_{k=0}^{\infty} \phi^k w_{t-k} \right) \right] \\
&= \sigma_w^2 \sum_{j=0}^{\infty} \phi^j \phi^{j+h} = \sigma_w^2 \phi^h \sum_{j=0}^{\infty} \phi^{2j} = \frac{\sigma_w^2 \phi^h}{1 - \phi^2}, \quad h \geq 0 \tag{3.78}
\end{aligned}$$

La varianza de un proceso autorregresivo mostrada en (3.79), se obtiene calculando la función de autocovarianza cuando $h = 0$. También se puede obtener tomando varianzas en $x_t = c + \phi x_{t-1} + w_t$.

$$\begin{aligned}
\sigma^2 &= \phi_1^2 \sigma^2 + \sigma_w^2 \\
\sigma_w^2 &= \sigma^2 (1 - \phi_1^2)
\end{aligned}$$

$$\sigma^2 = \frac{\sigma_w^2}{1 - \phi_1^2} \quad (3.79)$$

La función de autocorrelación se obtiene de la forma (3.80). Además $\rho(h)$ satisface la expresión $\rho(h) = \phi\rho(h - 1)$.

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)} = \frac{\frac{\sigma_w^2 \phi^h}{1 - \phi^2}}{\frac{\sigma_w^2}{1 - \phi_1^2}} = \phi^h \quad (3.80)$$

Finalmente, como se observa en la

Fig. 0.14, entre la serie y el rezago $h = 1$ existe una *asociación lineal significativa*. De acuerdo con el correlograma el *coeficiente de autocorrelación de orden uno* $\rho(1)$, entre la serie y el retraso $h = 1$ es de 0.6647. Sin embargo, $\rho(1)$ se debilita al aumentar el orden del retraso, ya que para $h = 2$ la autocorrelación es 0.4, y para $h = 3$, $\rho(3)$ no es significativo, mostrando que no existe autocorrelación para $h \geq 4$. Por lo tanto, estas observaciones se pueden modelar con un proceso autorregresivo de orden uno, $AR(1)$.

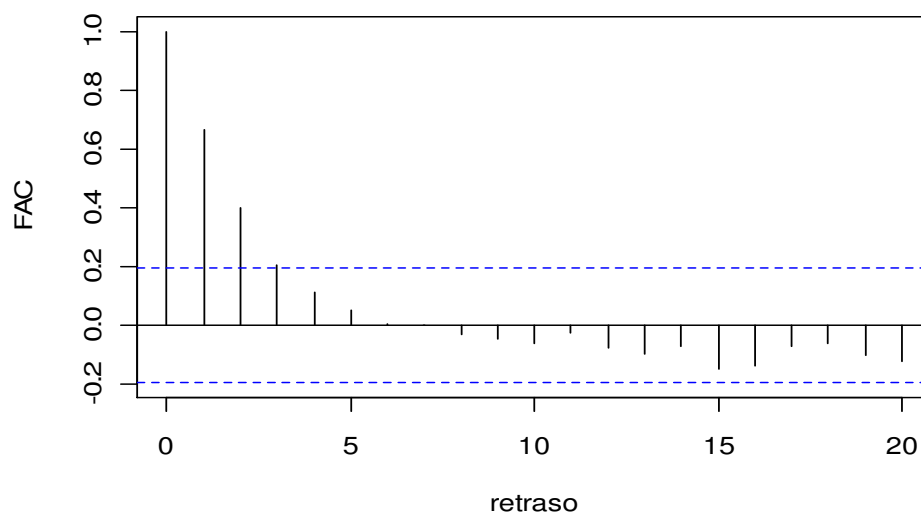


Fig. 0.14 Un Proceso $AR(1)$

CAPÍTULO 4. MATERIALES Y MÉTODOS

La metodología del proyecto de investigación se estructura con una secuencia de cinco etapas ilustradas en la Fig. 4.1. De manera general, en la etapa uno se determina las consecuencias de tratar datos autocorrelacionados con métodos clásicos. Posteriormente en las etapas dos y tres se trata la construcción de las cartas clásicas para datos independientes y se evalúa su desempeño mediante su ARL_0 . Para obtener el ARL_0 de la carta clásica se consideran datos sin autocorrelación y con autocorrelación. Finalmente en la cuarta y quinta etapa, se hace una modificación a las cartas clásicas, con el fin de mejorar su capacidad de tratar datos con y sin correlación de una manera eficiente. A continuación se hace una descripción detallada de cada etapa.

Etapa 1.- Efectos de ignorar la Correlación Serial en la Estimación de la Desviación Estándar.

Uno de los principales inconvenientes de utilizar cartas clásicas para tratar datos autocorrelacionados, es que por los métodos clásicos se obtiene una estimación deficiente de la desviación estándar real del proceso. En esta etapa se compara la *desviación estándar del proceso $AR(1)$* obtenida por los métodos del rango y la desviación estándar con la *desviación estándar real del proceso $AR(1)$* . Esta comparación se hace con la finalidad de cuantificar la diferencia entre la estimación hecha por los métodos tradicionales y el valor real de la desviación estándar correspondiente a un proceso con autocorrelación.

Etapa 2.- Construcción Teórica de las Cartas Clásicas: Caso Independiente.

En esta etapa se trabaja con tres cartas clásicas: carta para medias, varianzas y rangos. De acuerdo con la literatura, ha existido un gran interés en monitorear la centralidad de un proceso y varios autores han discutido las propiedades de las cartas de control para medias en presencia de autocorrelación tales como: *Alwan y Roberts (1988)* y *Runger y Willemain (1996)*. No obstante, de acuerdo con *Amin y Schmid (1997)* y *Kiani y Psarakis (2008)* no se ha realizado un estudio profundo de las cartas para dispersión cuando los datos son autocorrelacionados.

Inicialmente para la segunda etapa, se revisan los métodos necesarios para estimar los parámetros de las cartas que permitan la colocación adecuada de los límites de control. Así mismo, se obtienen los valores de las constantes tales como $d_2(n)$ y $d_3(n)$, las cuales están en función del tamaño de la muestra y se requieren específicamente en el *método del rango* para estimar la desviación estándar y para la construcción de la *carta de rangos*. Por otro lado la constante $c_4(n)$ es requerida para estimar la desviación estándar del proceso por el *método de la desviación estándar*.

Etapa 3.- Evaluación del ARL de las Cartas Clásicas para datos con y sin Autocorrelación.

En cuanto a la etapa número tres, se estudia el procedimiento de obtención del *ARL en control* para las *cartas de medias, varianzas y rangos* para diferentes tamaños de muestra. Este procedimiento involucra obtener el número promedio de muestras hasta detectar una falsa alarma dentro del proceso. De manera más específica, en cada *carta clásica* se obtiene el ARL_0 para datos *sin autocorrelación* y luego para datos con un *parámetro de*

autocorrelación mayor que cero, desde 0.1 hasta 0.9. Los datos utilizados se modelan mediante un proceso $AR(1)$. Finalmente, se evalúa cuál carta de dispersión resulta más afectada en términos del ARL_0 con el fin de hacer la modificación pertinente únicamente a una carta para monitorear la dispersión y a otra para la centralidad del proceso.

Etapa 4.- Construcción Teórica de las Cartas Modificadas y evaluación de ARL_0 para Caso Independiente y de Autocorrelación.

En cuanto a la etapa número cuatro, se procede a utilizar un método apropiado para estimar la *desviación estándar y varianza real* del proceso $AR(1)$. Con la modificación que se realiza sobre las cartas clásicas, ahora se tiene una carta para monitorear la centralidad y otra para la dispersión con límites de control adecuados. Se evalúa el ARL_0 para diferentes valores del parámetro de autocorrelación, de manera que la carta modificada es capaz de tratar datos con y sin autocorrelación arrojando un ARL_0 estable.

Etapa 5.- Comparación del Desempeño de Cartas Clásicas y Modificadas.

Finalmente, la etapa número cinco consiste en una discusión acerca del comportamiento que presenta el *ARL en control* para diferentes valores del parámetro de autocorrelación y para cambios en la desviación estándar del proceso. De manera general, en esta etapa se observa que el ARL_0 permanece estable ya que la carta modificada tiene la capacidad de tratar datos autocorrelacionados.

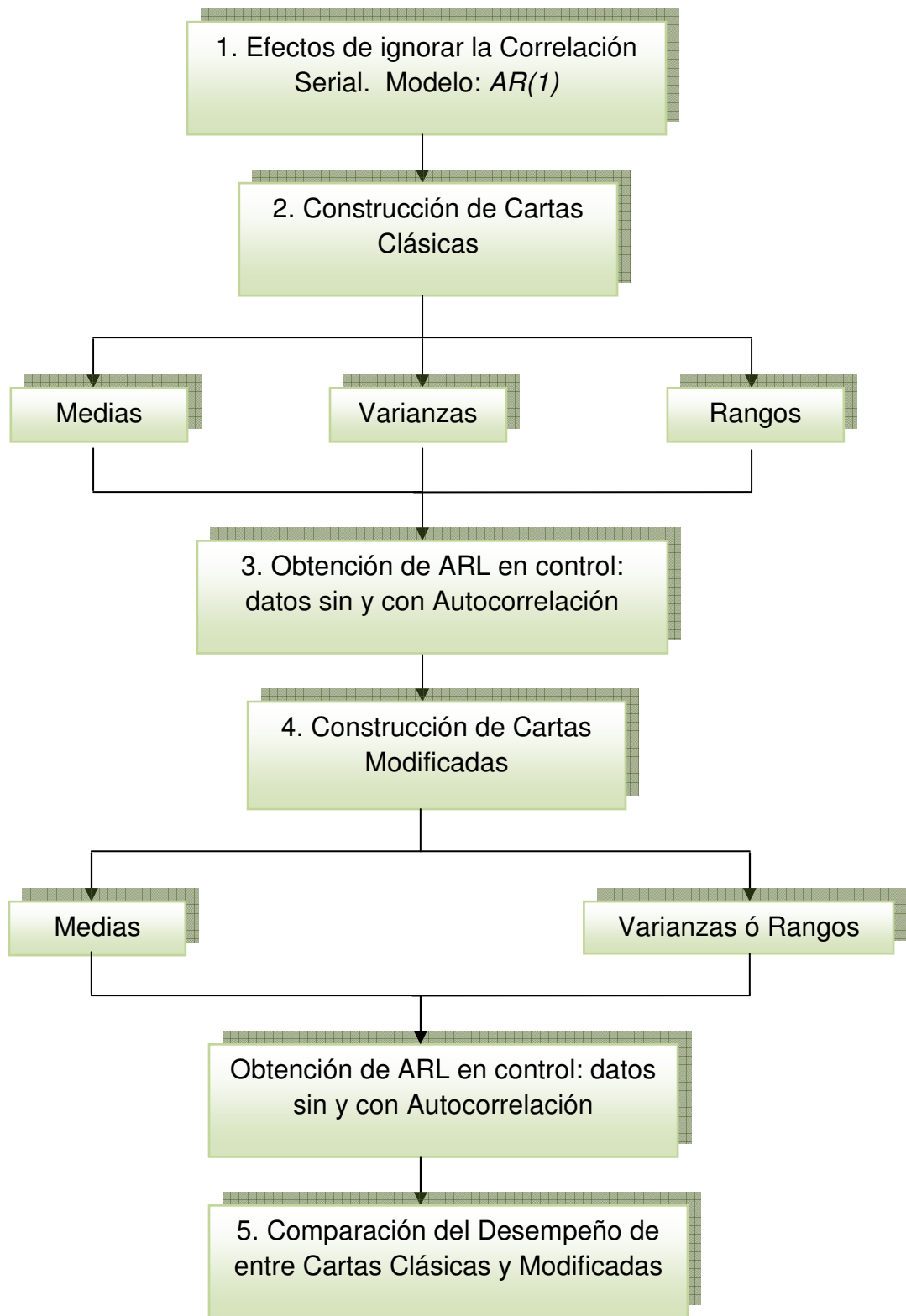


Fig. 0.1 Metodología del Proyecto de Investigación

La metodología del caso de estudio (Fig. 0.2) se estructura como se indica a continuación:

1.- Recolección de información del Proceso con el fin de elaborar una Base Histórica de Datos.

El proceso *fabricación del rotor automotriz* es realizado mediante cuatro operaciones: *Corte de ficha, Barrenado, Fosfatizado y Embutido*. Esta investigación está enfocada a los procesos de conformado de metales, por lo que únicamente se tratan datos provenientes del Corte de Ficha y Embutición. Los datos se recolectan en tres formas: por lotes, por subgrupos y en forma consecutiva.

2.- Realizar Pruebas de Autocorrelación a los datos mediante el Software R.

De acuerdo con *Shumway y Stoffer (2006)*, se requiere ajustar un modelo a las observaciones que exhiben autocorrelación, con el fin de tratar los datos adecuadamente. Es por eso, que las mediciones provenientes del proceso son analizadas mediante el software R, con el fin de encontrar un modelo que se ajuste al conjunto de observaciones. Además se determinan los parámetros del modelo, en el caso de un proceso *AR(1)*, se estima el coeficiente de autocorrelación para el retraso uno.

Para esta etapa se utilizan herramientas gráficas del software R, tales como el diagrama de dispersión y correlograma; así como la pruebas para autocorrelación: *Durbin- Watson*, y para estacionareidad: *Dickey-Fuller*.

3.- Si no existe Correlación Serial, sugerir el uso de las Cartas Clásicas. En caso contrario, utilizar las Cartas Modificadas.

Se sigue un procedimiento para verificar los supuestos de normalidad e independencia de las observaciones provenientes del proceso. Después de analizar la base histórica de datos, si se encuentra correlación serial o autocorrelación entre las observaciones, se propone usar las cartas modificadas para tratar adecuadamente esos datos. Es importante considerar si el proceso es o no estacionario, ya que la carta propuesta está diseñada para procesos estacionarios. Finalmente, para comprobar que la carta funciona adecuadamente se obtiene el ARL_0 .

Por otro lado, si los datos no siguen una distribución normal y se cuenta con una cantidad moderada de observaciones, *Vermaat (2006)* comenta que se pueden utilizar métodos Bayesianos para estimar los límites de control. Sin embargo, si se cuenta con una gran cantidad de datos, es mejor utilizar métodos no paramétricos.

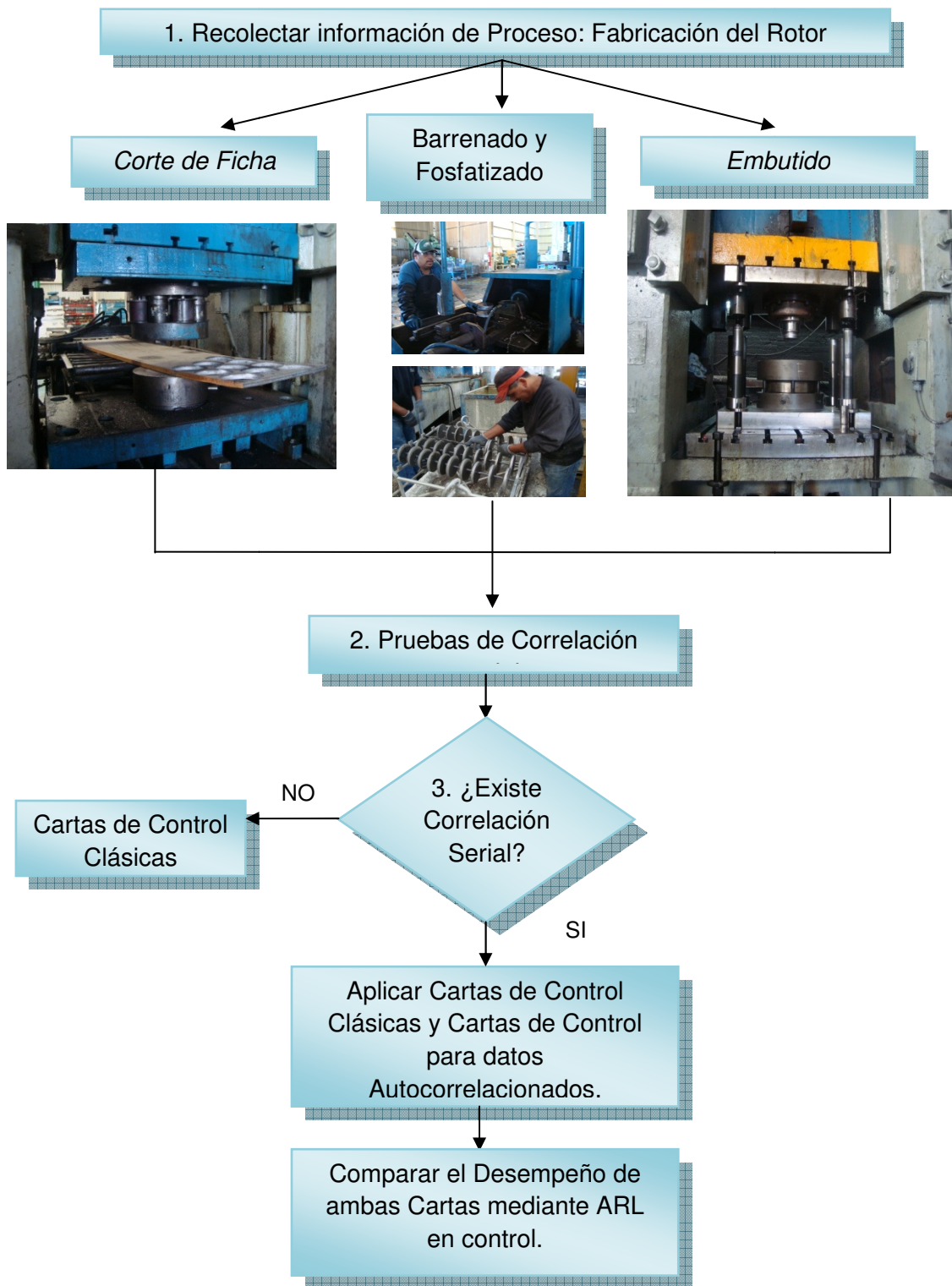


Fig. 0.2 Metodología: Caso de Estudio.

CAPÍTULO 5. ANÁLISIS DE RESULTADOS

En el presente capítulo se presenta la información correspondiente al caso de estudio desarrollado en la empresa automotriz Sanden Mexicana S.A. de C.V. Posteriormente se muestran y discuten los resultados obtenidos de la metodología descrita en el capítulo anterior.

Caso de Estudio.

En los procesos automatizados y de alta velocidad, frecuentemente se presenta autocorrelación en algunas características del producto. Esto se debe, en gran parte, a que la capacidad del proceso es dependiente del material y de la máquina, por lo que la autocorrelación es significativa. En particular, esto ocurre en el área de estampado de la empresa Sanden Mexicana S.A. de C.V. donde se presentan defectos en la fabricación del *rotor automotriz* tales como marcas, rebaba en contorno de la pieza, rayas y golpes. El proceso de fabricación del rotor se divide en las siguientes etapas (ver Fig. 0.1):

a) En el área de estampado se realiza el corte de ficha, para esto se utiliza materia prima en forma de hoja decapada (el proveedor aplica ácido clorhídrico y óxido superficial sobre la lámina para remover impurezas) con medidas de 400 x 1250 x 14 mm de espesor. Esto se lleva a cabo en una prensa mecánica de 1200 toneladas de capacidad.

b) Diámetro interior de la ficha barrenado en un torno horizontal, el barreno sirve para que posteriormente la pieza se fije a la matriz de la prensa donde se realiza el embutido.

c) Fosfatizado, consiste en un recubrimiento que actúa como un lubricante de fosfato de zinc para evitar la fricción entre el herramental y el producto.

d) Embutido, se fija la pieza a la matriz y se lleva a cabo el embutido de un solo paso. La prensa que se utiliza en esta operación es de 1200 toneladas.

e) En el área de rotor la pieza pasa por tres operaciones de maquinado y un ranurado. Después se ensambla con otros componentes que también son fabricados por Sanden Mexicana S.A. de C.V.

f) El ensamble se envía a Sanden Internacional, donde se finaliza el producto "Compresor de aire acondicionado".

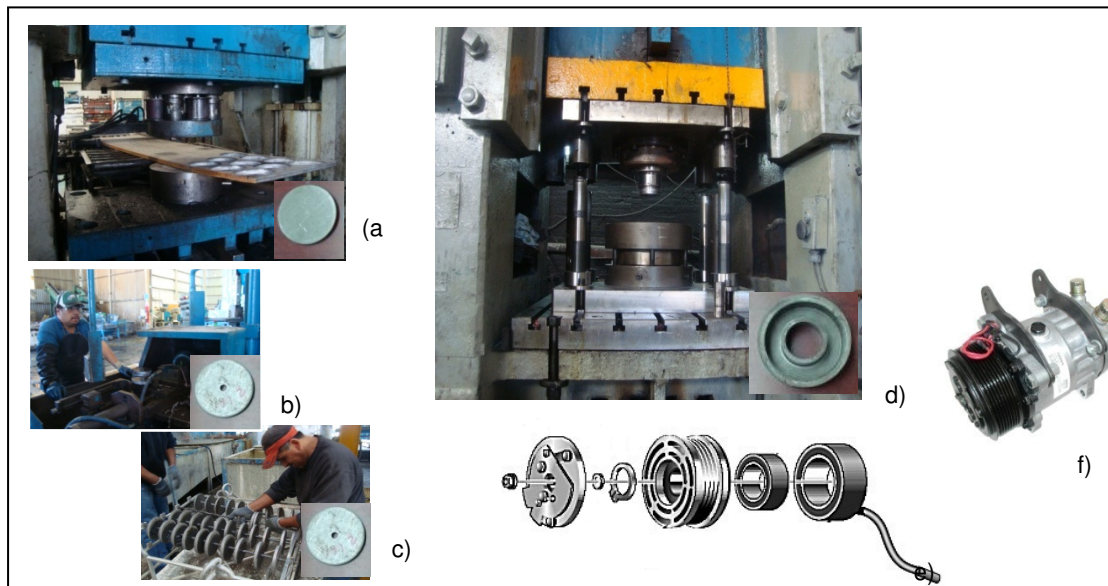


Fig. 0.1 Proceso de Fabricación de Rotor Automotriz.

Considerando que la operación de embutido es la más crítica, se realizó un diagnóstico del problema con 100 mediciones de las características espesor y altura total. La frecuencia de muestreo fue de cinco piezas cada diez minutos.

Para obtener el valor de los coeficientes de autocorrelación entre las observaciones, se construye la gráfica de función de autocorrelación o correlograma para las variables. En un correlograma se grafican la estimaciones de la función de autocorrelación muestral mediante la ecuación 5.1 (Montgomery, 2005). El numerador corresponde a la covarianza entre una variable y ella misma desfasada o retrasada en k lapsos de tiempo. El denominador corresponde a la varianza de las observaciones.

$$r_k = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t-k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2} \quad k = 0, 1, \dots, K \quad (5.1)$$

En la Fig. 0.2 para la variable *espesor* se tiene que en el retraso o desfase número uno la autocorrelación es aproximadamente 0.6, para el resto de los retrasos hasta el número veinte la autocorrelación es significativamente diferente de cero ya que los valores están por encima de los límites.

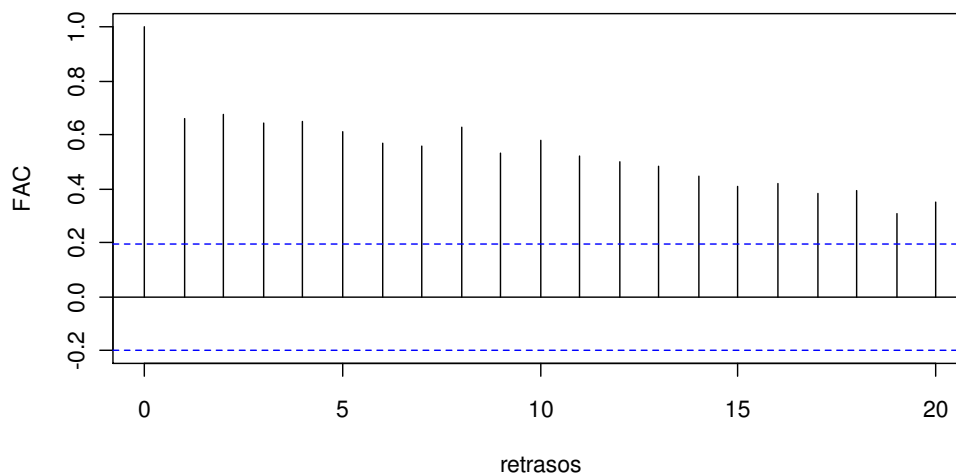


Fig. 0.2 Correlograma de la variable *Espesor*.

En la Fig. 0.3 para la variable *espesor* se presentan gráficas de dispersión entre la variable y su retraso k , donde $k = 1, 2 \dots 9$. Estas gráficas representan la autocorrelación entre la variable y sus retrasos. De manera que si se nombra a la variable *espesor* como x , se supone que x_1, x_2, \dots, x_n son observaciones de la variable. Representar la gráfica de dispersión correspondiente al retraso $k = 1$ requiere sustraer la primera observación x_1 y graficar el resto de las observaciones en el eje vertical. Para el eje horizontal se procede a sustraer la última observación x_n del resto de las observaciones. Para el retraso $k = 2$, se sustraen las primeras dos observaciones y las últimas dos y así respectivamente para cada retraso. De esta manera se presenta una gráfica de dispersión para la variable y su respectivo retraso k .

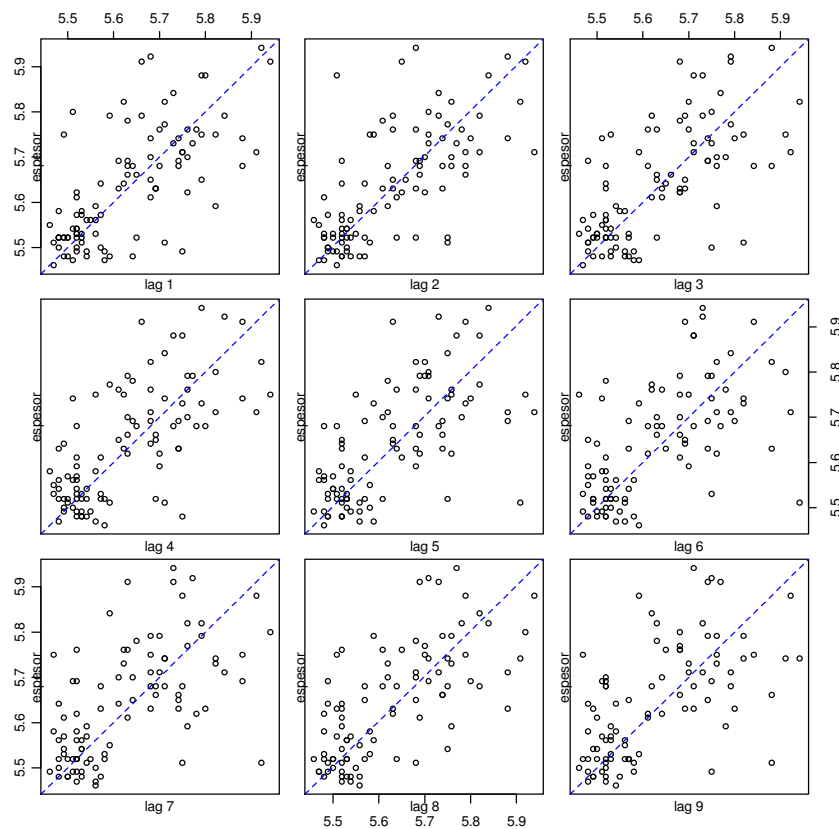


Fig. 0.3 Gráficas de Dispersión para la variable Espesor y sus retrasos.

En la Fig. 0.4 para la variable *altura total* se tiene que en los retrasos tres y siete la autocorrelación es ligeramente diferente de cero, para el resto de los retrasos hasta el número veinte la autocorrelación es muy cercana a cero o poco significativa.

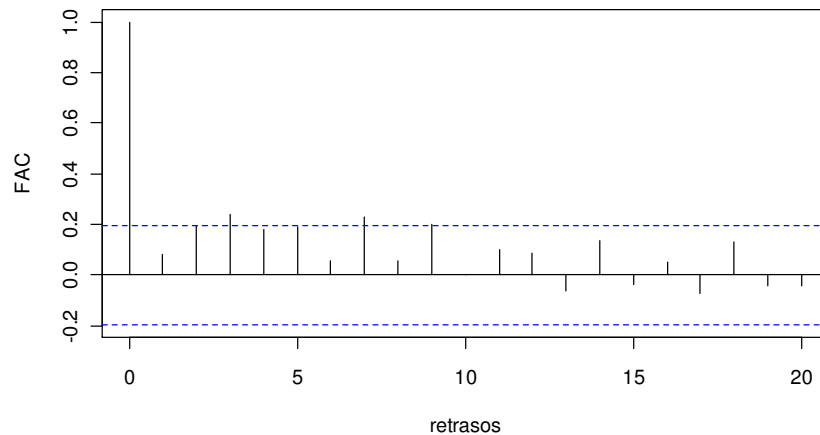
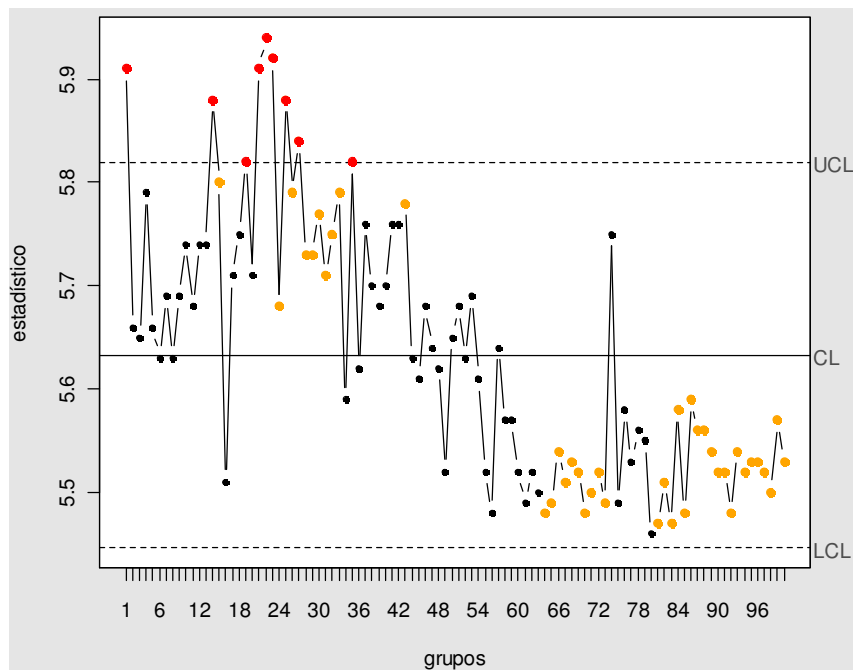


Fig. 0.4 Correlograma de la variable Altura total.

Con este diagnóstico se observa que en el proceso de fabricación del rotor automotriz se presenta correlación serial positiva de orden superior para la variable *espesor* y de orden inferior para la variable *altura total*. La correlación serial en este tipo de procesos se debe a que son altamente dependientes de la maquinaria y las condiciones de la materia prima, (AIAG *Statistical process control, 2005*).

Es en este contexto, en el cual uno de los principales supuestos para la metodología clásica de las cartas de control no se cumple. Las observaciones autocorrelacionadas son dependientes y para una carta de control clásica es necesario que los datos tomados del proceso sean independientes. Bajo estas condiciones, las cartas de control clásicas no son confiables y es necesario utilizar cartas de control que tomen en cuenta datos serialmente correlacionados.

Para observar el desempeño de las cartas de control cuando los datos son autocorrelacionados se grafican las observaciones de la variable *espesor* en una carta de control para observaciones individuales. Como se muestra en la Fig. 0.5 la carta presenta una gran cantidad de alarmas (puntos marcados en color rojo) así como violación de reglas de corridas (puntos marcados en color amarillo). Debido al patrón no aleatorio que se muestra en los datos y a los análisis realizados anteriormente, es evidente la presencia de correlación serial positiva lo que ocasiona la gran cantidad de alarmas señaladas en la carta.



Metodología Propuesta.

5.1 Etapa 1: Efectos de ignorar la Correlación Serial

Cuando un proceso es representado con un *modelo AR(1)* y las observaciones son graficadas en una *carta de control clásica*, el desempeño de la carta decrece ya que la autocorrelación afecta la estimación que se hace sobre la varianza y desviación estándar del proceso.

Para ilustrar lo anterior, en las Tabla 0.1 y Tabla 0.2 se presentan los efectos en la estimación de la desviación estándar del proceso ocasionados por ignorar la correlación serial positiva y negativa respectivamente; con diferentes valores del parámetro de autocorrelación, ρ , de un proceso *AR(1)*. En la segunda columna se estima la *desviación estándar real* del proceso *AR(1)*, σ_Y , y en columnas tres y cuatro se estima la *desviación estándar* del mismo proceso utilizando los *métodos tradicionales de rangos* y de *desviación estándar* con tamaños de muestra $n = 4$ y $n = 5$, respectivamente. En las tablas se muestra que σ_Y es subestimada para $\rho > 0$ y sobreestimada para $\rho < 0$, empleando los métodos tradicionales de estimación de la desviación estándar.

Tabla 0.1 Efectos de ignorar la Correlación Serial Positiva.

ρ	σ_Y	$E[\bar{R}/d_2]$		$E[\bar{S}/c_4]$	
		$n=4$	$n=5$	$n=4$	$n=5$
0.2677	1.0570	0.9984	0.9998	1.0171	1.0176
0.5071	1.0810	0.9223	0.9600	0.9212	0.9629
0.7526	1.1120	0.9196	0.9510	0.9116	0.9571

0.9471	1.1400	0.8948	0.8916	0.8970	0.9167
--------	--------	--------	--------	--------	--------

Tabla 0.2 Efectos de ignorar la Correlación Serial Negativa.

ρ	σ_Y	$E[\bar{R}/d_2]$		$E[\bar{S}/c_4]$	
		$n=4$	$n=5$	$n=4$	$n=5$
-0.3788	1.0830	1.1616	1.1550	1.1676	1.1719
-0.5679	1.1840	1.3837	1.3852	1.3876	1.4020
-0.7044	1.2930	1.6257	1.6139	1.6877	1.6653
-0.9438	1.4050	3.3228	3.0635	3.7510	3.5315

Lo anterior deja en claro que los *métodos tradicionales para estimar la desviación estándar* (ver capítulo 3 para una descripción detallada), no son apropiados cuando los datos provienen del proceso son autocorrelacionados. Utilizando estos métodos se obtiene una *estimación sesgada* de la desviación estándar del proceso, ya que las observaciones provienen de un proceso $AR(1)$. Por consecuencia, los límites de control de la carta no se calculan adecuadamente.

5.2 Etapa 2: Construcción Teórica de Cartas de Control Clásicas

En este documento, se estudian únicamente cartas de variables para controlar la *centralidad y dispersión* del proceso. Es importante recordar que las *cartas clásicas de Shewhart tradicionales* fueron diseñadas para observaciones que son normales e independientes. A continuación se menciona cada una de las *cartas clásicas* utilizadas para la parte experimental del trabajo.

Carta para Medias

La *carta clásica de Shewhart para medias*, \bar{X} , emplea una secuencia de promedios ordenada en el tiempo, \bar{X}_i , tomada de los subgrupos racionales. Para construir los límites de control de la carta de medias y revisar las estimaciones necesarias cuando no se conocen los parámetros poblacionales, véase la sección 3.2.3.

Carta para Varianzas

La *carta de control para varianzas*, S^2 , tiene el objetivo de monitorear cambios en la *varianza* del proceso. De acuerdo con *Khoo y Sim (2005)*, esta carta es eficiente cuando se requiere detectar cambios grandes en la *desviación estándar* del proceso. Como se mostró en la sección 3.2.3., construir la carta para varianzas requiere límites de control probabilísticos, ya que el estadístico $(n - 1)S^2/\sigma^2$ sigue una distribución ji-cuadrada con $n - 1$ grados de libertad, χ_{n-1}^2 .

Carta para Rangos

La distribución normal tiene un parámetro de centralidad y otro de dispersión, media μ y varianza σ^2 , respectivamente. Debido a lo anterior, no es suficiente controlar la media de un proceso de producción, (*Di Bucchianico, 2008*). Una manera sencilla de monitorear la *dispersión* de un proceso es la *carta para rangos*, ya que los cálculos para construir esta carta requieren

menos tiempo que los necesarios para construir otras *cartas de dispersión*, tales como cartas para varianzas o desviaciones estándar.

Como se mencionó en la sección 3.2.3., se requiere calcular los valores de las constantes $d_2(n)$ y $d_3(n)$ para estimar el *valor esperado* y la *desviación estándar del rango*, así como la *desviación estándar del proceso*. En la Tabla 0.3 se muestran los resultados para $E(R^2)$, que corresponde a la doble integral presentada en la ecuación (3.26), (ver anexo para encontrar el código introducido en Wolfram Mathematica). Así mismo, en la columna dos se presentan los valores para la constante $d_2(n)$. De acuerdo con (3.25), para obtener el valor de $d_3(n)$ se requiere restar $E^2(R_i)$ mostrado en columna tres, de la cantidad $E(R_i^2)$ ubicada en columna uno.

Tabla 0.3 Constantes d_2 y d_3 para diferentes Tamaños de Muestra.

n	$E(R^2)$	$d_2(n)$	$(d_2(n))^2$	$(d_3(n))^2$	$\sqrt{(d_3(n))^2}$
2	2.0000	1.1283	1.2732	0.7267	0.8525
3	3.6539	1.6925	2.8647	0.7892	0.8883
4	5.0125	2.0587	4.2384	0.7740	0.8798
5	6.1565	2.3259	5.4099	0.7466	0.8640
6	7.1424	2.5344	6.4232	0.7191	0.8480
7	8.0077	2.7043	7.3135	0.6942	0.8332
8	8.7786	2.8472	8.1065	0.6721	0.8198
9	9.4736	2.9700	8.8210	0.6525	0.8078
10	10.1063	3.0775	9.4710	0.6352	0.7970

Para una revisión más detallada de la construcción de la carta para rangos examine la sección 3.2.3., donde se especifican los límites de control

para cuando la *desviación estándar*, σ , es conocida; y para el caso donde se desconoce el valor de σ .

5.3 Etapa 3: Evaluar el *ARL* para las Cartas Clásicas.

En la *fase I* de una carta de control, es fundamental remover las causas especiales de variación y las muestras correspondientes; con el fin de que las muestras restantes provean información concerniente al estado en control del proceso. Así mismo, para la *fase II* es esencial contar con una tasa pequeña de falsas alarmas y la habilidad de detectar rápidamente puntos fuera de control. Una carta de control es una forma particular de una *prueba de hipótesis ordinaria*, no obstante se distingue de ésta por el hecho de que las cartas de control en la fase II involucran *pruebas secuenciales*. Esto es, en cada momento que se cuenta con nuevo subgrupo racional, se verifica si el proceso se encuentra dentro de control, H_0 , o fuera de control, H_1 . En este sentido, la *probabilidad de falsa alarma* en la carta de control está relacionada al *error tipo I* ó *error α* , de la prueba de hipótesis. En cambio, la habilidad de *detectar cambios* en el proceso está relacionada al *error tipo II* ó *error β* , de la prueba.

Para incorporar el carácter secuencial de las cartas de control en la Fase II, las nociones de error tipo I y II son reemplazadas por las propiedades de la *longitud de corrida*, definida como el número de subgrupos racionales observados hasta que la carta de control señala una alarma. En la literatura, la media es la propiedad de la longitud de corrida estudiada más profundamente. Por lo anterior, se utiliza el criterio de *ARL* como un *indicador de desempeño* para las siguientes dos situaciones: cuando el proceso se encuentra en control y cuando no lo está. Para esto, se emplean dos tipos de *ARL*: el *ARL* en control que se denota como ARL_0 , y el *ARL* fuera de control como ARL_1 . Obviamente, el ARL_0 debe ser grande (tasa reducida de falsas alarmas o error tipo I pequeño) y el ARL_1 debe ser pequeño (rápida detección de cambios y error tipo

El grande). El ARL_0 se define como el número de muestras promedio hasta que se detecta una falsa alarma y el ARL_1 como el número de muestras promedio recolectadas hasta detectar un cambio de tamaño δ . De manera adicional, el término ATS (Tiempo promedio para detectar), de acuerdo a sus siglas en inglés; se refiere al tiempo promedio transcurrido hasta que se presenta una señal. Para obtener el ATS , se multiplica el ARL por la frecuencia de muestreo expresada en horas.

Por otro lado, si se dispone de una secuencia de observaciones independientes, $\{X_t\}$, tales observaciones pueden corresponder a la media, varianza, rango o desviación estándar de una muestra o a una observación individual. La varianza de las observaciones, σ_X^2 , se asume como constante. De manera que, una observación de una característica de calidad en el tiempo t se representa:

$$X_t \sim \mathcal{N}(\mu_t, \sigma_X^2) \quad \text{para } t \in \mathbb{Z} \quad (5.2)$$

Tomando en cuenta lo anterior, una causa especial de variación en la carta de control puede ocurrir en un punto del tiempo desconocido T , tal que:

$$\mu_t = \begin{cases} \mu & \text{para } t < T \\ \mu + \delta\sigma_X & \text{para } t \geq T \end{cases} \quad (5.3)$$

donde el tamaño de los cambios provocados por una causa especial de variación es expresado en unidades de desviación estándar de las observaciones y se denota como $\delta\sigma_X$. Después de que un cambio ocurre, se puede escribir $P(\delta)$, la probabilidad de que una observación esté dentro de los límites de control:

$$P(\delta) = P(\mu - 3\sigma_X \leq X_t \leq \mu + 3\sigma_X)$$

$$= \Phi(\delta + 3) - \Phi(\delta - 3) \quad (5.4)$$

donde $\Phi(\cdot)$ denota la función de distribución acumulada de la distribución normal estándar. Dado que $P(0) = 0.9973$, es muy improbable observar una señal fuera de control si $\delta = 0$, es decir si no se presenta un cambio en la media del proceso.

De acuerdo con *Wieringa (1999)*, el *ARL* se define como el número promedio de observaciones hasta e incluyendo la primera observación fuera de control. De manera que, el *ARL* para observaciones independientes se calcula como sigue:

$$ARL = \frac{1}{P(\cdot)} \quad (5.5)$$

Si el proceso se encuentra en control, $P(\cdot)$ es α y el ARL_0 se obtiene de la forma: $ARL = 1/\alpha$. Por el contrario, si el proceso se encuentra fuera de control, $P(\cdot)$ es $1 - \beta$ y el ARL_1 se obtiene de la forma: $ARL = 1/(1 - \beta)$.

Carta para Medias.

En una carta clásica de medias para calcular el ARL_0 , es decir cuando $\delta = 0$; y el ARL_1 , cuando $\delta > 0$; se requiere obtener $P(\delta)$ tal como se expresa a continuación:

$$P(\delta) = \Phi(3 + \delta\sqrt{n}) - \Phi(-3 - \delta\sqrt{n}) \quad (5.6)$$

donde el término δ expresa el tamaño del cambio en la media, medido en unidades de desviación estándar y n se refiere al tamaño de muestra. En la

Tabla 0.4 se muestra el ARL_0 para diferentes valores de δ y n . Se observa que el ARL_0 es alto si $\delta = 0$, y que el ARL_0 decrece si δ es grande. Por otra parte, cuando una carta de control con el mismo $ARL(0)$ y un $ARL(\delta)$ más bajo para $\delta > 0$ es más eficiente, es decir, en promedio, pocas observaciones o muestras son necesarias para detectar un cambio en la media del proceso. Por el contrario, una carta de control con el mismo $ARL(0)$ y un $ARL(\delta)$ más alto para $\delta > 0$, es menos eficiente o sensible; ya que se requiere recolectar una gran cantidad de observaciones o muestras para detectar un cambio en el parámetro que se está monitoreando.

Tabla 0.4 ARL para Carta de Medias Clásica, con diferentes valores de δ y n .

δ	n								
	2	3	4	5	6	7	8	9	10
0	370.398	370.398	370.398	370.398	370.398	370.398	370.398	370.398	370.398
0.5	90.6462	60.6879	43.8946	33.4007	26.3575	21.3826	17.7308	14.9676	12.8251
1	17.7308	9.7647	6.3029	4.4953	3.4366	2.7656	2.3154	2.0000	1.7716
1.5	5.2690	2.9080	2.0000	1.5664	1.3334	1.1995	1.1198	1.0715	1.0423
2	2.3154	1.4733	1.1885	1.0758	1.0296	1.0110	1.0039	1.0013	1.0004
2.5	1.4207	1.1010	1.0232	1.0048	1.0008	1.0001	1.0000	1.0000	1.0000
3	1.1198	1.0142	1.0013	1.0001	1.0000	1.0000	1.0000	1.0000	1.0000

De manera general en la carta para medias, se tiene que la probabilidad de una falsa alarma cuando el proceso está bajo control, es igual a:

$$\begin{aligned}
 1 - P(LIC < \bar{X}_i < LSC | \mu = \mu_0) &= \phi(-3) + (1 - \phi(3)) = 2(1 - \phi(3)) \\
 &\approx 0.00270 \approx \frac{1}{370} \quad (5.7)
 \end{aligned}$$

donde la probabilidad $P(0)$ corresponde a 0.0027, en base a esto se asegura que 0.27% de las observaciones se encuentran fuera de los límites y el resto, es decir 99.73% están dentro de los límites de control. Por lo tanto,

una falsa alarma se presenta en promedio cada 370 observaciones o muestras recolectadas. Estos porcentajes disminuyen cuando δ es mayor que cero.

Con el fin de estudiar el comportamiento de la carta de control clásica se obtiene su ARL_0 mediante simulación, para el caso de datos autocorrelacionados. Para esto se estima la media y la desviación estándar del proceso con la información proveniente del caso de estudio. Luego se determinan los límites de control de la forma clásica, posteriormente se generan datos mediante la función *arima.sim* en el software R. Los datos corresponden a un proceso $AR(1)$, con diferentes valores para el parámetro de autocorrelación ρ y se comparan contra los límites de control para obtener el ARL_0 de la carta.

Se observa en Tabla 0.5 y Tabla 0.6 para tamaños de muestra $n = 4$ y $n = 5$, que cuando $\rho = 0$ y $\delta = 0$, es decir cuando no hay correlación ni cambios en la media, el ARL_0 permanece estable con un valor de 370. En cambio, a medida que el parámetro ρ se incrementa, el ARL_0 decrece. Por lo tanto, es evidente que la correlación serial afecta el desempeño de la carta.

A medida que n aumenta y con las condiciones $\rho > 0$ y $\delta > 0$, el deterioro en el desempeño de la carta es más evidente, mostrándose en la disminución del ARL_0 . Este inconveniente ocurre debido a la subestimación de la desviación estándar, tal como se vio en la Etapa 1 de la experimentación; ocasionando que los límites se hagan estrechos y que la cantidad de falsas alarmas aumente.

Tabla 0.5 ARL para Carta de Medias Clásica, $n=4$.

ρ	0σ	0.5σ	1σ	1.5σ	2σ	2.5σ	3σ
--------	-----------	-------------	-----------	-------------	-----------	-------------	-----------

0	369.4734	43.9699	6.2942	2.0074	1.1879	1.0236	1.0012
0.1	183.2991	31.2384	5.6057	2.0089	1.2170	1.0332	1.0026
0.2	90.2209	21.8006	4.9858	1.9817	1.2489	1.0464	1.0057
0.3	45.3661	15.7235	4.4657	1.9945	1.2828	1.0681	1.0108
0.4	24.6863	11.2216	4.0617	1.9955	1.3298	1.0936	1.0200
0.5	14.1000	8.2387	3.6353	1.9957	1.3787	1.1313	1.0361
0.6	8.3166	5.9510	3.2465	1.9948	1.4289	1.1749	1.0635
0.7	5.4232	4.2654	2.8934	1.9714	1.4908	1.2367	1.1058
0.8	3.2005	2.9774	2.4364	1.9214	1.5605	1.3197	1.1794
0.9	2.1438	1.9960	1.8800	1.7258	1.5602	1.4092	1.2873

Tabla 0.6 *ARL* para Carta de Medias Clásica, $n=5$.

ρ	0σ	0.5σ	1σ	1.5σ	2σ	2.5σ	3σ
0	371.7242	34.2178	4.5089	1.5659	1.0751	1.0049	1.0001
0.1	175.8784	23.6651	4.0854	1.5963	1.0943	1.0086	1.0002
0.2	82.0688	17.1108	3.7975	1.6204	1.1231	1.0151	1.0007
0.3	41.0592	12.5542	3.5144	1.6531	1.1516	1.0262	1.0025
0.4	22.3442	9.2970	3.2477	1.6893	1.1911	1.0435	1.0064
0.5	12.2026	6.8181	3.0512	1.7193	1.2425	1.0700	1.0157
0.6	7.3749	5.0699	2.7921	1.7588	1.3028	1.1132	1.0331
0.7	4.4852	3.7280	2.5275	1.7734	1.3688	1.1679	1.0689
0.8	2.8725	2.6776	2.1884	1.7627	1.4513	1.2571	1.1362
0.9	1.8982	1.8639	1.7595	1.6105	1.4583	1.3412	1.2441

Cartas para Rangos.

En una carta para rangos se grafica el rango, R , de cada subgrupo racional que se recolecta. Tal como se mostró en el capítulo 3, para muestras aleatorias de tamaño n , $E(R_i/d_2) = \sigma$. Por lo tanto, σ se puede estimar mediante $\hat{\sigma} = \bar{R}/d_2$. Además la desviación estándar de R , σ_R , puede estimarse utilizando $\hat{\sigma}_R = d_3\bar{R}/d_2$, donde las constantes d_2 y d_3 se presentan en la sección 5.2.

El ARL para la carta de rangos se obtiene mediante (5.8) y para el caso independiente, la probabilidad $P(R > LSC(R))$ es como se expresa en (5.9).

$$ARL(R) = \frac{1}{P(R > LSC(R))} \quad (5.8)$$

$$\begin{aligned} P(R > LSC(R)) &= 1 - F_R(LSC) \\ &= 1 - \int_{-\infty}^{\infty} [\phi(z + d_2(n) + 3 * d_3(n) - \phi(z))^{n-1} \phi(z) * dz \quad (5.9) \end{aligned}$$

donde $\phi(\cdot)$ es la *función de distribución normal estándar* y $\phi(\cdot)$ es la *función de densidad de la distribución normal estándar*. En la Tabla 0.7 se observa el ARL_0 obtenido para diferentes tamaños de muestra, n , y para diferentes valores del parámetro de autocorrelación, ρ . Los resultados se resumen a continuación:

1. En el caso de $n = 2$ y $n = 3$, el ARL_0 se incrementa hasta alcanzar valores muy elevados, para valores de ρ grandes.

2. Para $5 \leq n \leq 10$, el ARL_0 se incrementa y luego decrece para valores pequeños de ρ , sin embargo disminuye tan pronto como ρ es grande.
3. Para $\rho \in [0.4, 0.9]$ el ARL_0 decrece a medida que n aumenta de tamaño.

Tabla 0.7 ARL para Carta de Rangos Clásica.

ρ/n	2	3	4	5	6	7	8	9	10
0	203.458	204.057	204.115	206.402	201.454	203.424	202.052	201.438	203.176
0.1	159.030	244.763	235.680	253.065	233.176	242.648	208.985	240.688	259.930
0.2	224.191	323.672	299.080	299.758	236.086	240.083	235.712	241.165	202.351
0.3	331.702	362.193	332.733	259.005	219.485	203.812	215.740	178.915	168.946
0.4	515.659	375.690	275.367	228.785	180.471	163.270	145.603	142.521	132.135
0.5	683.571	415.678	282.079	188.433	167.145	120.735	102.719	96.223	80.026
0.6	964.894	447.308	247.822	157.222	113.358	82.184	68.139	58.868	48.252
0.7	1219.961	448.940	224.123	123.088	82.184	59.341	44.213	36.297	30.509
0.8	1409.717	464.162	181.633	97.452	58.363	40.531	29.454	22.461	18.456
0.9	1738.505	477.111	157.150	70.834	41.385	27.322	19.402	14.495	11.602

Es evidente que para un proceso autorregresivo de orden uno, $AR(1)$, la correlación serial positiva ocasiona que la carta de rangos tenga un desempeño deficiente, que se ve reflejado en la inflación del ARL_0 para $n = 2$ y $n = 3$ y en la disminución del ARL_0 para $n \geq 4$ y valores de ρ elevados.

Carta para Varianzas.

De acuerdo con *Khoo y Sim (2005)*, la *varianza muestral* tiene una distribución asimétrica por lo que no se recomienda una carta de control basada en ± 3 desviaciones estándar como la mayoría de las cartas de Shewhart. Debido a lo anterior, los límites de control de la carta clásica para varianzas se construyen a partir de que la cantidad $(n - 1)S^2/\sigma^2$ sigue una distribución ji-cuadrada, $\chi_{\alpha, n-1}^2$. El ARL de la carta de varianza se calcula mediante la siguiente ecuación:

$$ARL(S_i^2) = 1/P(S_i^2 > \hat{\sigma}^2 \chi_{\alpha, n-1}^2 / (n - 1)) \quad (5.10)$$

donde $\chi_{\alpha, n-1}^2$ denota el valor crítico de la distribución ji-cuadrada con $n - 1$ grados de libertad y $\hat{\sigma}^2$ es la estimación de la *varianza* obtenida de una base de datos. Dependiendo del valor de α que se desee fijar, el *ARL* estable, es decir ARL_0 , resulta en diferentes valores. Para $\alpha = 0.002$, el ARL_0 resulta en 500; en cambio cuando $\alpha = 0.005$, se obtiene un ARL_0 de 200. En la Tabla 0.8 se muestran los valores críticos para diferentes α .

Tabla 0.8 Valores Críticos de la Distribución ji-cuadrada.

g.d./ α	0.001	0.002	0.004	0.005	0.008	0.01
1	10.8275	9.5495	8.2838	7.8794	7.0334	6.6348
2	13.8155	12.4292	11.0429	10.5966	9.6566	9.2103
3	16.2662	14.7955	13.3164	12.8381	11.8269	11.3448
4	18.4668	16.9237	15.3656	14.8602	13.7889	13.2767
5	20.5150	18.9073	17.2789	16.7496	15.6251	15.0862
6	22.4577	20.7911	19.0987	18.5475	17.3748	16.8118
7	24.3218	22.6006	20.8491	20.2777	19.0604	18.4753
8	26.1244	24.3520	22.5451	21.9549	20.6961	20.0902
9	27.8771	26.0564	24.1973	23.5893	22.2913	21.6659
10	29.5883	27.7216	25.8130	25.1881	23.8531	23.2092

En el estudio de simulación se utiliza únicamente el valor de 0.005 para α , de modo que se pretende obtener un ARL_0 de 200. Los resultados se muestran en la Tabla 0.9 para diferentes valores de ρ y n . La información se resume a continuación:

1. Cuando $\rho = 0$, es decir cuando los datos son independientes, el ARL_0 permanece estable.
2. En el caso de $n = 2$ y $n = 3$, el ARL_0 se incrementa hasta alcanzar valores muy elevados para valores de ρ grandes.

3. Para $4 \leq n \leq 10$, el ARL_0 se incrementa y luego decrece para valores pequeños de ρ , sin embargo disminuye tan pronto como ρ es grande.
4. Para $\rho \in [0.2, 0.9]$ el ARL_0 decrece a medida que n aumenta de tamaño.

Tabla 0.9 ARL para Carta de Varianza Clásica.

ρ/n	2	3	4	5	6	7	8	9	10
0	204.975	200.272	197.242	194.417	205.156	205.824	196.192	203.139	197.522
0.1	296.333	274.376	238.962	223.853	216.495	209.812	230.349	259.815	276.330
0.2	558.797	357.112	242.789	221.935	174.247	148.975	175.055	172.509	145.813
0.3	862.789	388.700	263.606	178.480	167.246	147.023	105.979	131.358	103.288
0.4	856.132	514.587	297.243	156.971	126.043	98.636	83.961	61.319	71.646
0.5	1689.520	490.811	279.850	137.890	88.246	61.263	51.396	43.519	38.977
0.6	2208.541	670.091	232.728	134.735	64.813	50.194	29.786	27.650	23.349
0.7	4003.065	731.187	186.313	76.958	45.748	32.772	20.881	17.070	14.274
0.8	4327.531	842.221	166.798	56.377	31.854	21.292	15.376	10.879	9.662
0.9	7181.787	708.156	125.800	52.428	25.312	15.931	11.331	8.636	6.549

La carta de varianzas presenta una disminución del ARL_0 para $n \geq 5$ y $\rho > 0$, superior al que presenta la carta de rangos. Además, el ARL_0 de la carta de varianza presenta una mayor inflación para $n = 2$, $n = 3$ y $\rho > 0$, que el ARL_0 de la carta de rangos. Por lo tanto, es indiscutible que la presencia de autocorrelación causa efectos más negativos en la carta de varianzas que en la de rangos.

5.4 Etapa 4: Construcción de Cartas Modificadas y su ARL.

Una carta modificada es una extensión de los métodos clásicos de CEP (exclusivos para observaciones normales e independientes) que consiste en estimar la varianza de la serie de tiempo en lugar de estimar la varianza suponiendo que los datos no exhiben ningún tipo de autocorrelación. Como se explicó en la sección anterior, la *carta de varianzas* resulta ser la *carta clásica para dispersión* más afectada por la autocorrelación. Por lo tanto, en los siguientes párrafos se explica únicamente la modificación aplicada a las cartas de medias y varianzas.

Carta para Medias.

Se dispone de observaciones que corresponden a un modelo $AR(1)$ expresado de la forma:

$$Y_{i,j} = \mu_0 + \rho(Y_{i,j-1} - \mu_0) + \varepsilon_{i,j} \quad (5.11)$$

donde el parámetro de autocorrelación es $\rho < 1$. Las variables aleatorias $\varepsilon_{i,j}$ son independientes y distribuidas normalmente con valor esperado 0 y varianza σ_0^2 . Por lo tanto $E(Y_{i,j}) = \mu_0$ y varianza, $\gamma_v = \sigma_0^2 \rho^{|v|} / (1 - \rho^2)$. El modelo expresado en (5.11), se basa en que las observaciones entre muestras o subgrupos racionales son independientes. Por el contrario, las observaciones dentro de una muestra si están correlacionadas. Para la construcción de la carta modificada de medias se estima la varianza del proceso $AR(1)$ de la forma:

$$\sigma^2 = \frac{\sigma_w^2 \rho^j}{1 - \rho^2} \quad (5.12)$$

Se aclara que cuando un proceso autorregresivo es de orden 1, entonces j toma el valor de uno y con la ecuación (5.12) se obtiene la autocovarianza para el retraso $j = 1$. Sin embargo, si se desea obtener la varianza del proceso que tiene datos autocorrelacionados, se fija j con un valor de cero. Cuando se dispone de subgrupos racionales de tamaño n , la estimación de la *varianza del proceso con datos autocorrelacionados* se obtiene de la forma:

$$Var_{\rho}(\bar{X}) = \frac{1}{n} * \frac{\sigma_0^2}{1 - \rho^2} \left[1 + 2 \sum_{j=1}^{n-1} \left(1 - \frac{j}{n} \right) \rho^j \right] \quad (5.13)$$

donde el parámetro de autocorrelación se denota como ρ y el retraso j es igual a uno. En base a esto se define la regla de decisión para la carta modificada de medias:

$$|\bar{X}_i - \mu_0| > x_m \sqrt{Var_{\rho}(\bar{X})} \quad (5.14)$$

para alguna constante x_m , el índice m hace referencia a la carta modificada. Con el cálculo de la varianza real del proceso, los límites modificados para la carta son ahora adecuados. De acuerdo a las Tabla 0.10 y Tabla 0.11, las cartas modificadas tienen la capacidad de trabajar con datos serialmente correlacionados, ya que se puede observar un desempeño estable para el ARL_0 en cada valor de ρ .

Tabla 0.10 *ARL* para Carta Modificada de Medias, $n=4$.

ρ	0σ	0.5σ	1σ	1.5σ	2σ	2.5σ	3σ
0	370.3983	43.8946	6.3029	2.0000	1.1885	1.0232	1.0013
0.1	369.7594	44.7153	6.2411	2.0026	1.1922	1.0237	1.0015
0.2	371.7487	44.4039	6.3436	2.0207	1.1915	1.0243	1.0009
0.3	367.2152	44.3559	6.3662	1.9849	1.1947	1.0226	1.0014
0.4	368.6975	42.9111	6.2950	2.0116	1.1839	1.0228	1.0014
0.5	369.6891	43.7350	6.2241	2.0099	1.1880	1.0241	1.0017
0.6	371.7464	42.6740	6.2985	2.0027	1.1890	1.0222	1.0012
0.7	370.7302	44.6668	6.2916	1.9789	1.1865	1.0225	1.0012
0.8	372.9072	44.7938	6.2690	2.0010	1.1924	1.0245	1.0012
0.9	369.8546	44.3713	6.3133	1.9984	1.1860	1.0230	1.0013

Tabla 0.11 *ARL* para Carta Modificada de Medias, $n=5$.

ρ	0σ	0.5σ	1σ	1.5σ	2σ	2.5σ	3σ
0	368.2103	32.5751	4.4644	1.5747	1.0771	1.0051	1.0001
0.1	369.8072	32.9690	4.5128	1.5600	1.0737	1.0052	1.0001
0.2	370.5074	34.3449	4.5719	1.5607	1.0770	1.0043	1.0001
0.3	371.8805	34.1884	4.5008	1.5639	1.0767	1.0048	1.0001
0.4	370.6375	32.6700	4.4365	1.5721	1.0745	1.0049	1.0001
0.5	369.6818	32.8997	4.4897	1.5749	1.0782	1.0045	1.0001
0.6	369.4113	33.5014	4.5319	1.5648	1.0777	1.0044	1.0001
0.7	368.3674	34.0767	4.5062	1.5649	1.0751	1.0043	1.0000
0.8	372.0481	34.3045	4.4850	1.5664	1.0752	1.0052	1.0000

0.9	370.1097	34.3334	4.4280	1.5655	1.0760	1.0045	1.0000
-----	----------	---------	--------	--------	--------	--------	--------

Carta para Varianzas

Se dispone de muestras que provienen de un modelo autorregresivo $AR(1)$ y se calcula la varianza para cada muestra de tamaño n , de la forma:

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (X_{i,j} - \bar{X}_i)^2 \quad (5.15)$$

El índice i hace referencia al número de muestra que se recolecta, entonces la regla de decisión para la carta de varianza es:

$$S_i^2 > Var_{\rho}(S^2) \chi_{\alpha, n-1}^2 / (n-1) \quad (5.16)$$

donde el valor esperado del estadístico, $E_{\rho}(S^2)$, se muestra en la ecuación (5.17); y la varianza, $V_{\rho}(S^2)$, en la ecuación (5.18).

$$E_{\rho}(S^2) = \gamma_0 \left\{ 1 - \frac{2}{n-1} \sum_{j=1}^n \left(1 - \frac{j}{n}\right) \rho^j \right\} \quad (5.17)$$

$$V_{\rho}(S^2) = \frac{2}{(n-1)^2} \gamma_0^2 \left\{ \sum_{i=1}^n \sum_{j=1}^n \rho^{2|i-j|} - \frac{2}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \rho^{|i-j|+|i-k|} + \frac{1}{n^2} \left(\sum_{i=1}^n \sum_{j=1}^n \rho^{|i-j|} \right)^2 \right\} \quad (5.18)$$

el término γ_0 es la varianza del proceso que tiene datos autocorrelacionados. Debido a que no se dispone de un valor numérico para la

varianza, ésta debe estimarse de la base de datos del proceso. Ya que se cuenta con muestras preliminares, se estima γ_0 de la forma:

$$\hat{\gamma}_0 = \frac{1}{m} \sum_{i=1}^m S_i^2 \quad (5.19)$$

En la Tabla 0.12 se observa la estimación de la varianza, γ_0 , para $n = 2$ y $n = 5$, asumiendo que los datos son independientes aunque en realidad provienen de un proceso $AR(1)$. Con respecto a la columna tres, se presenta la estimación de la varianza real del proceso, $Var_{\rho}(S^2)$, considerando el parámetro de autocorrelación, ρ .

Tabla 0.12 Estimación de Varianza para un Proceso $AR(1)$.

ρ	$n = 2$		ρ	$n = 5$	
	γ_0	$Var_{\rho}(S^2)$		γ_0	$Var_{\rho}(S^2)$
0.1	1.0951	1.7740	0.1	1.0047	0.4647
0.2	0.7802	0.9986	0.2	1.0125	0.4326
0.3	0.9639	0.9446	0.3	0.9665	0.3772
0.4	0.9188	0.6615	0.4	0.9987	0.3476
0.5	0.8850	0.4425	0.5	1.0600	0.3154
0.6	0.5573	0.1783	0.6	1.0117	0.2398
0.7	0.7297	0.1313	0.7	0.9639	0.1614
0.8	0.6175	0.0494	0.8	0.9349	0.0879
0.9	0.5228	0.0104	0.9	0.9202	0.0275

En la Tabla 0.13 se obtiene el ARL_0 de la carta modificada de varianzas para diferentes tamaños de n y ρ . Se observa que el ARL_0 permanece estable para $\rho \geq 0$, lo que indica que la carta tiene un desempeño adecuado incluso cuando los datos son autocorrelacionados.

Tabla 0.13 *ARL* para Carta Modificada de Varianzas.

ρ/n	2	3	4	5	6	7	8	9	10
0	199.936	199.487	200.284	199.756	200.434	201.517	200.049	201.301	201.616
0.1	197.170	201.430	203.150	200.752	199.050	200.054	201.067	201.062	200.063
0.2	201.390	198.955	200.856	199.920	199.158	201.390	201.646	200.198	200.607
0.3	201.276	202.234	198.043	200.925	199.638	201.021	199.476	199.933	200.847
0.4	199.866	199.426	197.134	198.417	200.439	199.995	199.220	200.503	199.090
0.5	201.188	201.643	196.213	200.752	199.507	199.215	201.266	200.314	200.533
0.6	200.740	197.140	204.473	199.314	198.199	201.025	200.800	201.446	200.866
0.7	199.303	200.951	196.251	199.768	199.916	198.740	201.098	200.535	199.112
0.8	199.375	201.506	199.645	199.652	200.979	201.243	201.681	199.549	200.292
0.9	199.908	201.229	197.535	201.958	201.895	199.269	201.550	201.638	201.247

5.5 Etapa 5: Comparaciones entre Cartas Clásicas y Modificadas.

En una carta clásica para medias, tal como se observó en la Tabla 0.4 para $\delta > 0$ el ARL_0 decrece en diferentes cantidades dependiendo del tamaño de muestra, n , como se muestra en la Fig. 0.6. La línea punteada en color azul corresponde a la curva de ARL_0 para $n = 2$, mientras que las líneas en amarillo, morado y rojo corresponden a $n = 3, 4$ y 5 ; respectivamente.

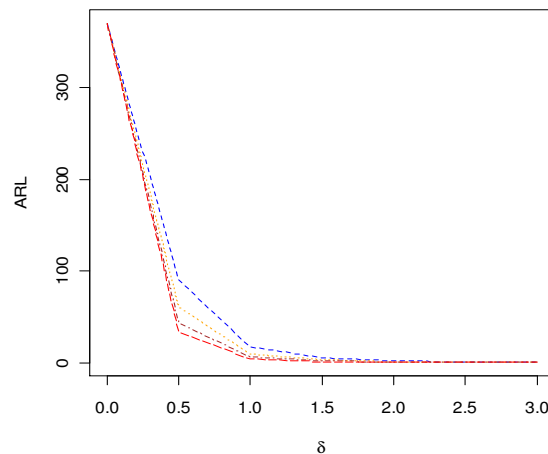


Fig. 0.6 Curva de *ARL*: Carta de Medias, $n=2,3,4$ y 5 .

Conforme el tamaño de muestra aumenta desde $n = 2, 3, \dots 5$ y $\delta > 0$, el ARL_0 disminuye cada vez más; tal que el ARL_0 para $\delta = 0.5$ y $n = 2$ es 90.6462, mientras que para $\delta = 0.5$ y $n = 3$ es 60.6879 y finalmente para $n = 5$ es 33.4007. Es notable que la curva disminuye conforme n aumenta y $\delta > 0$. De la misma forma el ARL_0 es afectado cuando $\rho > 0$, es decir cuando se presenta una estructura de autocorrelación.

A continuación, el logaritmo del ARL_0 , $\log(ARL_0)$, se grafica para diferentes cambios en la media representados por la constante δ . La línea punteada en azul representa la curva de la carta modificada y la línea en rojo la curva de la carta clásica. En las siguientes figuras (5.7-5.10) se compara el ARL_0 de ambas cartas para diferentes valores del parámetro de autocorrelación. Es claro que la carta modificada para medias tiene un mejor desempeño que la carta clásica para $\rho > 0$. De igual forma, ese comportamiento se repite para la carta modificada de varianzas. Fig. 0.7

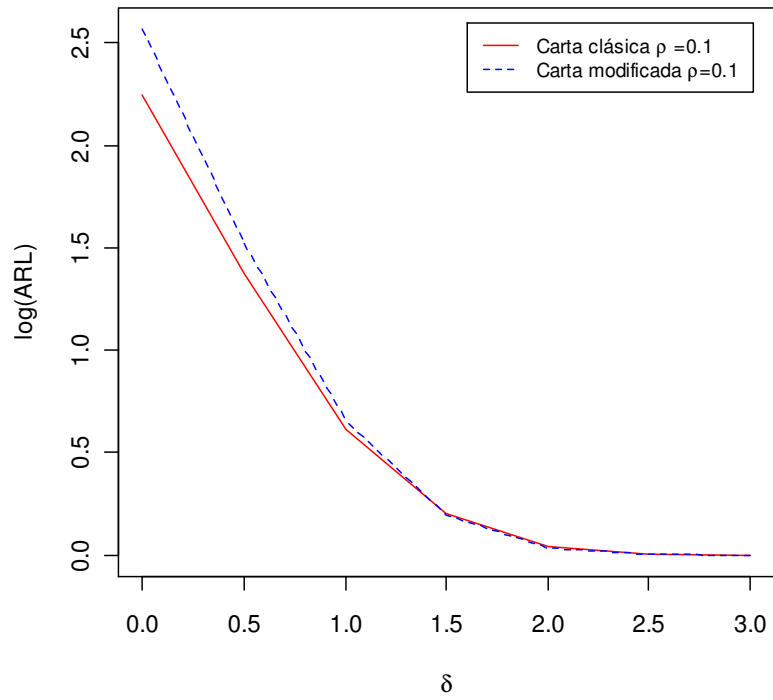


Fig. 0.7 Comparación Curvas de *ARL*: Carta Clásica y Modificada, $\rho=0.1$

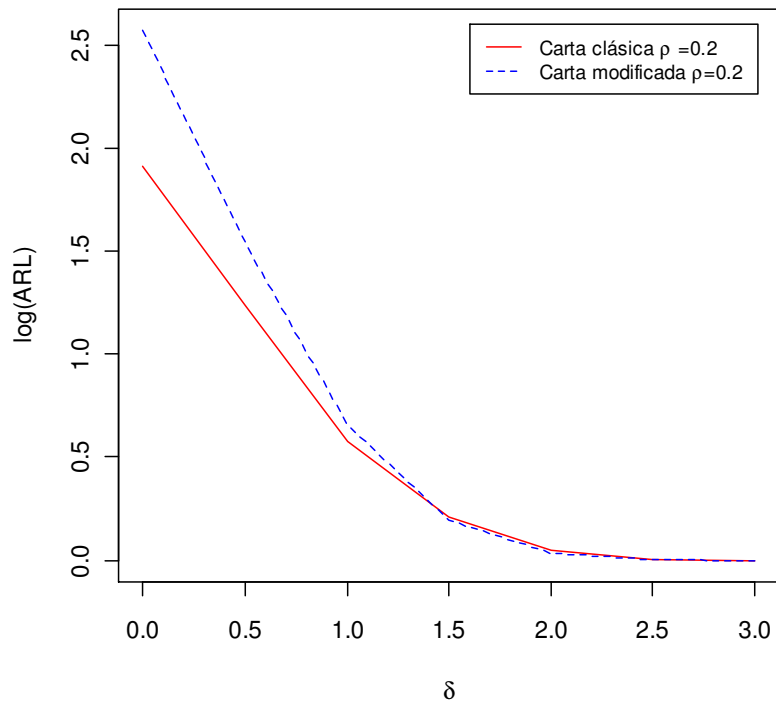


Fig. 0.8 Comparación Curvas de *ARL*: Carta Clásica y Modificada, $\rho=0.2$

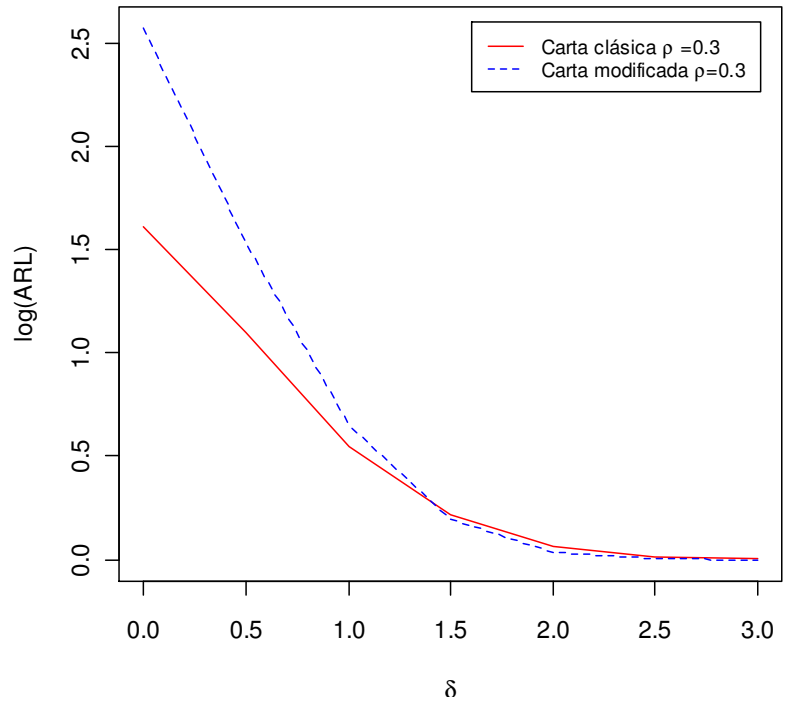


Fig. 0.9 Comparación Curvas de *ARL*: Carta Clásica y Modificada, $\rho=0.3$

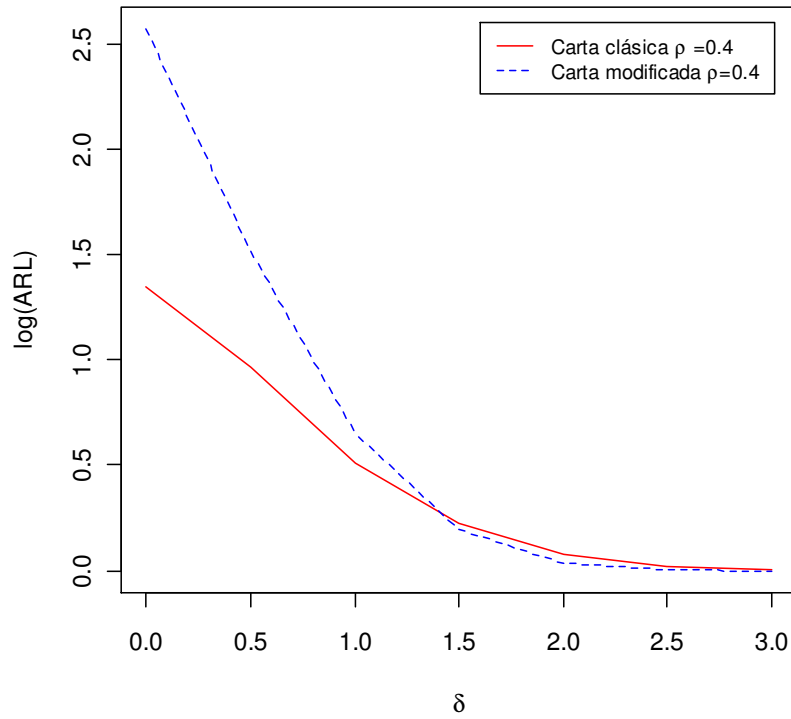


Fig. 0.10 Comparación Curvas de *ARL*: Carta Clásica y Modificada, $\rho=0.4$

CAPÍTULO 6. CONCLUSIONES Y RECOMENDACIONES

Validación de las Hipótesis

Se estudiaron los efectos de ignorar la *correlación serial* en las cartas clásicas de medias, varianzas y rangos. Desde el enfoque de la detección de un solo patrón: *puntos fuera de los límites*, la presencia de autocorrelación provoca que los límites de control se amplíen o se reduzcan innecesariamente, disminuyendo y aumentando el *error tipo I* respectivamente, y de forma opuesta para el *error tipo II*, como se muestra en la Tabla 0.1.

Tabla 0.1 Efectos de la Autocorrelación sobre los Límites de Control.

<i>Ampliando los límites de control</i>	<i>Disminuir los límites de control</i>
Decrece el <i>error tipo I</i> (riesgo de que un punto caiga fuera de los límites de control, indicando una condición fuera de control cuando no hay causas especiales).	Aumenta el error tipo I.
Incrementa el <i>error tipo II</i> (el riesgo de que un punto caiga dentro de los límites cuando el proceso está fuera de control).	Disminuye el error tipo II

De acuerdo a la sección 5.4, la *carta para varianzas* resultó ser más afectada por la correlación serial que la *carta de rangos*, en términos del ARL_0 . Considerando que la autocorrelación ocasionó una estimación errónea de la varianza del proceso, se requirió una modificación para las cartas, con el objetivo de que tuvieran la capacidad de tratar datos autocorrelacionados.

Una vez construidas las cartas modificadas de medias y varianzas, se demostró que el *desempeño* de las mismas es más eficiente, en términos del *ARL en control*, en comparación con las cartas clásicas. En base a la hipótesis alterna, conviene aclarar que en las cartas clásicas de rangos y varianzas para $n = 2, 3$ y 4 ; el *ARL en control* se eleva hasta valores muy grandes, volviéndose muy inestable. Así mismo para $n \geq 5$, el *ARL en control* toma valores por debajo del valor objetivo, hasta llegar a valores muy pequeños como 11.602 para el caso de rangos y 6.549 para varianzas. La hipótesis alterna queda probada, considerando que el ARL_{δ} de la carta modificada es estable y por lo tanto, la carta tiene una tasa baja de falsas alarmas.

Alcance de los Objetivos planteados y Recomendaciones

Se obtuvo una metodología de monitorización eficiente para los procesos que exhiben correlación serial. En cuanto al caso de estudio se encontró que las variables críticas del proceso, *espesor* y *altura total*, presentaron un comportamiento autorregresivo. Se trabajó con la base histórica de datos, utilizando la variable *espesor* y fue evidente que las cartas clásicas no tienen un desempeño adecuado. Posteriormente, a través de un estudio de simulación se obtuvo el ARL_0 de las cartas modificadas para diferentes valores del parámetro de autocorrelación, ρ , lo que mostró que estas cartas no presentan problemas ante datos autocorrelacionados.

De acuerdo con *Wieringa (1999)*, se debe evaluar de dos formas las causas que provocan la autocorrelación antes de interpretar y analizar la información:

1. Si el comportamiento autorregresivo es una parte normal e inamovible del proceso, entonces, utilizar una carta que detecta señales fuera de control a causa de la presencia de autocorrelación no es muy útil.

2. No obstante, si la autocorrelación es un síntoma de un problema que debe tratarse, se recomienda usar una carta de control que detecte la presencia de autocorrelación.

Con respecto al caso de estudio, las causas que provocan la autocorrelación en los datos del proceso corresponden a la situación número uno, donde la autocorrelación forma parte de las *causas comunes* y por lo tanto no se considera como una *causa especial* de variación. Anteriormente, las cartas de control en el proceso arrojaban señales fuera de control que en realidad eran *falsas alarmas*, ocasionadas por la autocorrelación positiva en los datos y no por alguna causa asignable de variación.

Trabajo Futuro

Como trabajo futuro se identifican tres áreas: *la estimación que se hace sobre la autocorrelación*, el caso de *variables que son multivariadas* y otro tipo de *cartas (EWMA o CUSUM)* para trabajar con datos provenientes de modelos tales como *ARMA o ARIMA*. En los siguientes párrafos se detalla cada uno de estos temas.

La estimación de la autocorrelación es una herramienta esencial de diagnóstico para identificar la estructura de dependencia en las series de tiempo y así mismo para modelar de una manera precisa el proceso. Sin embargo, de acuerdo con *Arnau y Bono (2001)*, la estimación de la

autocorrelación es sesgada para tamaños de muestra pequeños. El sesgo se incrementa considerablemente a medida que el tamaño de muestra se reduce.

Frecuentemente, las cartas de control son usadas para detectar cambios inusuales en variables que son independientes y que por lo tanto no son influenciadas por el comportamiento de otras variables. Sin embargo, las variables de un proceso a menudo están interrelacionadas y por lo tanto forman un proceso multivariado. De acuerdo con *Mason y Young (2002)*, dado que las variables no tienen un comportamiento independiente, deben ser examinadas como un grupo y no separadamente.

De acuerdo con *Borja y Páez (2008)*, utilizar la carta *EWMA* propuesta por *Montgomery y Mastrangelo*, en presencia de datos autocorrelacionados; requiere que el proceso sea estacionario y que las observaciones del proceso estén autocorrelacionadas en forma positiva, para que puedan representarse mediante un modelo de *Promedios Móviles Integrados Autorregresivos, ARIMA (0,1,1)*. Sin embargo, si esas condiciones no se cumplen, la carta no tendrá un buen funcionamiento. Considerando lo anterior, razones para investigaciones futuras en ese sentido pueden ser: a) desarrollar cartas alternativas donde el funcionamiento de las mismas no se vea afectado por el incumplimiento de los supuestos, b) usar otro modelo que se ajuste mejor en lugar del *ARIMA (0,1,1)* o c) utilizar cartas *MEWMA* para series de tiempo multivariadas.

BIBLIOGRAFÍA.

Libros, Folletos y Tesis.

- Arnold B.C., Balakrishnan N. y Nagaraja H.N. (2008). *A first course in Order Statistics*. Ed. John Wiley & Sons. Philadelphia.
- Bartés A. P., Llabrés X. T., Cintas P. G., Fernández L. P. y Vidal J. S, (2000). *Métodos Estadísticos: Control y mejora de la Calidad*. Editorial UPC (Universidad Politécnica de Catalunya). Barcelona.
- Guerrero V. M. (2003). *Análisis Estadístico de Series de tiempo Económicas*. 2da ed. Editorial Thomson.
- Kotz S. y Johnson N.L. (1993). *Process Capability Index*. Ed. Chapman & Hall/CRC. E.U.A.
- Mason R. L. y Young J. C. (2002). *Multivariate Statistical Process Control with Industrial Applications*. Ed. American Statistical Association and the Society for Industrial and Applied Mathematics. E.U.A.
- Montgomery D. C. y Runger G. C. (2007). *Probabilidad y Estadística aplicadas a la Ingeniería*. 3er ed. Ed. Limusa-Wiley.
- Montgomery D.C. (2009). *Introduction to Statistical Quality Control*. 6ta.ed. Ed. John Wiley & Sons.
- Moore S. D. (2005). *Estadística básica aplicada*. 2da.ed. Ed. Antoni Bosch S.A. Barcelona, España.
- Shumway R. H. y Stoffer D. S. (2006). *Time series Analysis and its applications with R examples*. 2da.ed. Ed. Springer Texts in Statistics. E.U.A.
- Vermaat M. B. (2006). *Statistical Control in non-standard situations*. Instituut Voor Bedrijfs- en Industriële Statistiek. Amsterdam.
- Wackerly D. D, Mendenhall W. y Scheaffer L. R. (2008). *Mathematical Statistics with applications*. 7a. ed. Ed. Duxbury.
- Wieringa J. E. (1999). *Statistical Process Control for Serially Correlated data*. Tesis doctoral. Ed. Labyrinth Publication.

Artículos de Revistas.

- Alwan L.C. y Roberts H.V. (1988). *Time-series Modeling for Statistical Process Control*. Journal of Business & Economic Statistics. Vol. 6, No.1, 87-95.
- Alwan L.C. y Roberts H.V. (1995). *The problem of Misplaced Control Limits*. Appl. Statist. Vol. 44, No.3, 269-278.
- Amin R.W., Schmid W. y Frank O. (1997). *The effects of Autocorrelation on the R-chart and the S² chart*. The Indian Journal of Statistics. Vol. 59, No. 2, 229-255.
- Apley D.W. (2002). *Time series Control Charts in the presence of Model Uncertainty*. Journal of Manufacturing Science and Engineering. Vol. 124. 891-898.
- Arnau J. y Bono R. (2001). *Autocorrelation and Bias in short Time series: An alternative Estimator*. Quality & Quantity Vol. 35, 365-387.
- Chang Y.M. y Wu T. L. (2009). *On Average Run Lengths of Control Charts for Autocorrelated Processes*. Methodology and Computing in Applied Probability. Vol. 13, No. 2, 419-431.
- Cheng S.W. y Thaga K. (2005). *Max-Cusum Chart for Autocorrelated Processes*. Statistica Sinica. Vol. 15, 527-546.
- Claro F.A., Costa A.F. y Machado M.A. (2008). *Double Sampling \bar{X} Control Chart for a first order Autoregressive Process*. Pesquisa Operacional. Vol. 28, No. 3, 545-562.
- Elevli S., Uzgören N. y Savas M. (2009). *Control Charts for Autocorrelated Colemanite data*. Journal of Scientific & Industrial Research. Vol. 68, 11-17.
- Ferrer A. (2004). *Control Estadístico de Procesos con dinámica: una revisión del Estado del Arte y Perspectivas del futuro*. Estadística Española, Vol. 46, Núm.155, 19-47.
- González B. J. y Páez P. J. (2008). *Estudio de la Carta EWMA en presencia de datos Autocorrelacionados*. Ciencia e Ingeniería. Vol. 2, No. 3, 11-25.
- Haridy S. y Wu Z. (2009). *Univariate and Multivariate Control Charts for monitoring Dynamic-Behavior Processes: a Case Study*. Journal of Industrial Engineering and Management. Vol. 2, No. 3, 464-498.

- Harris T. J. y Ross H. W. (1991). *Statistical Process Control Procedures for Correlated observations*. The Canadian Journal of Chemical Engineering. Vol. 69, No. 2, 48-57.
- Khoo M.B.C y Sim J.N. (2005). *Enhancements for the S^2 Control Chart*. Jabatan Matematik. Vol. 2, 133-141.
- Kiani M., Paraneetos J. y Psarakis S. (2008). *A new Procedure for monitoring the Range and Standard deviation of a quality characteristic*. Quality and Quantity. No. 9067.
- Knoth S. y Schmid W. (2002). *Monitoring the Mean and the Variance of a Stationary Process*. Statistica Neerlandica. Vol. 56, No. 1, 77-100.
- Kramer H. y Schmid W. (1998). *The influence of Parameter estimation on the ARL of Shewhart type Charts for Time series*. Statistical Papers. Vol. 41, 173-196.
- Lovelace C.R., Swain J.J., Zeinelabdin H. y Gupta J.N.D. (2009). *Lower Confidence Limits for Process Capability Indices C_p and C_{pk} when data are Autocorrelated*. Quality and Reliability Engineering International. Vol. 25, 663-700.
- Messaoud A., Weihs C. y Hering F. (2005). *Time series Analysis, Control Charts: an Industrial Application*. Fachbereich Statistik.
- Nembhard D.A. y Black H. (2000). *A Demerits Control Chart for Autocorrelated data*. Quality Engineering. Vol. 13, No. 2, 179-190.
- Noffsinger J. R. y Anderson R. B. (2002). *Effect of Autocorrelated data on Composite Panel Production Monitoring and Control: A comparison of SPC techniques*. Forest Products Journal. Vol. 52, No. 3, 60-79.
- Psarakis S. y Papaleonida G.E.A. (2007). *SPC Procedures for Monitoring Autocorrelated Processes*. Quality Technology & Quantitative Management. Vol. 4, No. 4, 501-540.
- Ramos A. W. y Ho L.L. (2003). *Inferential Procedures for Capability Indices in Autocorrelated data by Bootstrap*. Revista Produção. Vol. 13, No. 3.
- Runger G.C. y Willemain T.R. (1996). *Batch –means Control Charts for Autocorrelated data*. IIE Transactions. Vol. 28, No. 6, 483-487.
- Schmid W. (1995). *On the Run Length of a Shewhart Chart for Correlated data*. Statistical Papers. Vol. 36, 111-130.

Sun J., Wang S. y Fu Z. (2010). *Process Capability Analysis and Estimation scheme for Autocorrelated data*. J Syst Sci Syst Eng. Vol.19, No.1, 105-127.

Wardell D. G., Moskowitz H. y Plante R.D. (1992). *Control charts in the presence of data Correlation*. Management Science. Vol. 38, No. 8, 1084-1105.

Young M. T. y Winistorfer P. M. (2001). *The effect of Autocorrelation on real-time Statistical Process Control with solutions for Forest Products Manufacturers*. Forest Products Journal. Vol. 51, No. 11/12, 70-77.

Manuales y Cursos.

AIAG Statistical process control (SPC) (2005). *Manual reference. Second edition Daimler Chrysler Corporation, Ford Motor Company and General Motors Corporation*.

Di Bucchianico, A. (2008). *Lecture Notes 3TU Course Applied Statistics*, Technische Universiteit Eindhoven.

Juran, Frank M. Gryna, R.S. Bingham. (1990). *Manual de J.M.* Editorial Reverté.

ANEXOS

Códigos de programación en el paquete R. CAPITULO 3.

Fig. 3.2 Gráficas de puntos y dispersión con diferente desviación estándar.

```
par(mfrow=c(2,2))
set.seed(150)
datos=rnorm(50,5,0.7)
stripchart(datos,method="overplot",pch=1)
mtext(expression(sigma==0.7),side=1,line=4,font=1)
plot(datos, xlab= "Número de observación")
abline(h=5)
mtext(expression(sigma==0.7),side=1,line=4,font=0.5)
mean(datos)
var(datos)
sd(datos)

set.seed(150)
datos2=rnorm(50,5,0.25)
stripchart(datos2,method="overplot",pch=1, xlim=range(datos))
mtext(expression(sigma==0.25),side=1,line=4,font=1)
plot(datos2,ylim=range(datos),xlab= "Número de observación")
abline(h=5)
mtext(expression(sigma==0.25),side=1,line=4,font=0.5)
mean(datos2)
var(datos2)
sd(datos2)
```

Fig. 3.3 Gráficas de puntos y dispersión con diferente media.

```
par(mfrow=c(2,2))
set.seed(150)
datos=rnorm(50,5,0.25)
datos2=rnorm(50,5.25,0.25)

stripchart(datos,method="overplot",pch=1,xlim=c(4,6))
mtext(expression(mu==5),side=1,line=4,font=1)
plot(datos, ylim=c(4,6),xlab= "Número de observación")
abline(h=5)
mtext(expression(mu==5),side=1,line=4,font=0.5)
mean(datos)
var(datos)
sd(datos)

stripchart(datos2,method="overplot",pch=1, xlim=c(4,6))
mtext(expression(mu==5.25),side=1,line=4,font=1)
plot(datos2,xlab= "Número de observación",ylim=c(4,6))
abline(h=5.25)
mtext(expression(mu==5.25),side=1,line=4,font=0.5)
mean(datos2)
var(datos2)
sd(datos2)
```

Fig. 3.4 Comportamiento de $c_4(n)$ contra diferentes tamaños de muestra.

```
c4=c()
n = 2:50
for(i in 1:49){
c <- sqrt(2/(n[i]-1)) * gamma(n[i]/2)/gamma((n[i]-1)/2)
c4=c(c4,c)
}
c4
n=2:50
plot(n,c4, type="l")
r=c(n,c4)
dim(r) = c(length(n), 2)
r
```

Fig. 3.5 Comportamiento de $d_2(n)$ contra diferentes tamaños de muestra.

```
integrate(dnorm, -1.96, 1.96)
integrate(dnorm, -Inf, Inf)

d2 =c()
n <- 2:50
for(i in 1:49){
f <- function(y){ 1 - pnorm(y)^n[i] - (1 - pnorm(y))^n[i] }
d = integrate(f, -Inf, Inf)
d2 = c(d2,d$value)
}
d2
n=2:50
plot(n, d2, type="l")
m = c(n,d2)
dim(m) = c(length(n), 2)
m
```

Fig. 3.6 Tendencia ascendente en una serie de tiempo.

```
set.seed(150)
w = rnorm(200,0,1); x = cumsum(w)
wd = w +.2; xd = cumsum(wd)
plot.ts(xd, ylim=c(-5,55), xlab="Tiempo", ylab="Observaciones")
lines(.2*(1:200), lty="dashed")
```

Fig. 3.7 Variación cíclica en una serie de tiempo.

```
set.seed(149)
t = 1:500
c = 2*cos(2*pi*t/50 + .6*pi)
w = rnorm(500,0,1)
plot.ts(c + w, xlab="Tiempo", ylab="Observaciones")
```

Fig. 3.8 Proceso de ruido blanco gaussiano.

```
set.seed(150)
```

```
ww=rnorm(100,0,1)
plot.ts(ww, xlab="Tiempo", ylab="Observaciones")
```

Fig. 3.9 Proceso de promedios móviles.

```
set.seed(150)
w = rnorm(100,0,1) # 500 N(0,1)
v = filter(w, sides=2, rep(1,5)/5) # Promedio Móvil
plot.ts(v, ylim=c(-4,4), main="Promedios móviles de orden 5", xlab="Tiempo")
```

Fig. 3.10 Proceso autorregresivo.

```
set.seed(100)
w = rnorm(550,0,1) # 50 extra to avoid startup problems
x = filter(w, filter=c(1, -.9), method="recursive")
plot.ts(x[51:550], main="Un proceso autoregresivo", ylab="Observaciones", xlab="Tiempo")
abline(h=0)
```

Fig. 3.11 Caminata aleatoria.

```
set.seed(154)
w = rnorm(200,0,1); x = cumsum(w)
wd = w +.3; xd = cumsum(wd)
plot.ts(xd, ylim=c(-5,55), xlab="Tiempo", ylab="Observaciones")
lines(x)
lines(.3*(1:200), lty="dashed")
```

Fig. 3.12 Correlograma para una serie de tiempo.

```
set.seed(100)
Xt<-arima.sim(n=100, list(ar=c(0.9),ma=c(0)),sd=sqrt(0.25))
datos<-Xt+5
plot(datos, type="o", xlab="Número de observación", ylab="value", main="Un proceso con correlación positiva")
acf(datos, xlab="h", ylab="FAC")
```

Fig. 3.13 Un proceso con media constante.

```
set.seed(152)
plot(datos<-arima.sim(list(order=c(1,0,0),ar=.8),
n=100),type="o", xlab="Tiempo", ylab="Valores", main=(expression("AR(1) " "phi" = +.8, Un proceso con correlación positiva"))))
abline(h=(0))
```

Fig. 0.1 Un proceso AR(1)

```
set.seed(109)
Xt<-rnorm(100, 0,1)
x = filter(Xt, filter=c(0.9), method="recursive")
acf(x, xlab="retraso", ylab="FAC")
ar(x)
```

CAPITULO 5.

Fig. 5.2 Correlograma de variable Espesor.

```
datos=read.csv("datossanden.csv")
acf(datos$x2,main="FAC Espesor",xlab="retrasos",ylab="FAC")
```

Fig. 5.3 Gráficas de dispersión para la variable Espesor y sus retrasos.

```
datos=read.csv("datossanden.csv")
par(mfrow=c(2,1))
espesor=datos$x2
windows()
lag.plot(espesor,lags=9,diag=TRUE,diag.col="blue", do.lines=FALSE,cex.main=0.8)
```

Fig. 5.2 Correlograma de variable Altura total.

```
datos=read.csv("datossanden.csv")
acf(datos$x1,main="FAC Altura ", xlab="retrasos",ylab="FAC")
```

Fig. 5.5 Carta de control de medias para observaciones individuales de variable Espesor.

```
datos=read.csv("datossanden.csv")
espesor=datos$x2
require(qcc)
obj<-qcc(espesor, type="xbar.one")
```

Fig. 0.7-5.10 Comparación curvas de ARL: carta clásica y modificada, $\rho=0.1$

```
orange2=read.csv("arl1.csv")
orange2$x <- as.numeric(orange2$x)
nx<- max(orange2$x)

# Rango para ejes x y y.
xrange <- range(orange2$delta)
yrange <- range(orange2$logARL)

# Construir gráfica
plot(xrange, yrange, type="n", xlab=expression(" "delta" " ), ylab="log(ARL)" )
colors <- c("red", "blue")
linetype <- c(1:nx)

# Agregar líneas
for (i in 1:nx) {
  x <- subset(orange2, x==i)
  lines(x$delta, x$logARL, type="l", lwd=1.5,
  lty=linetype[i], col=colors[i])
}
x1=expression ("Carta modificada "rho" =0.1" )
x2=expression("Carta clásica "rho"=0.1 " )

# Agregar leyenda
legend(1.6, 2.6, c(x1, x2) , cex=0.8, col=colors, lty=linetype)
```

Tabla 0.1 y 5.2 Efectos de ignorar la correlación serial positiva y negativa.

Descargar paquete “qcc.groups”, y en la lista de la instrucción `arima.sim`, cambiar el parámetro de autocorrelación “ ρ ” en la parte “`ar`” del modelo, por un valor numérico positivo o negativo.

----MÉTODO TRADICIONAL DESVIACIÓN ESTÁNDAR, $n=4$ y $\phi=0.2677$ -----

```
X<-arima.sim(n=100, list(ar=c(0.2677),ma=c(0)),sd=1)
datos=X
sample<-gl(25,4)
subgrupos <- qcc.groups(datos, sample)
sd=c()
x=c()
for(i in 1:25){
x[i]=sd(subgrupos[i,])
}
sd=c(sd, x)
sdt=mean(sd)
```

-----Desviación estándar muestral-----

sdt

-----Valores para constante c_4 , $n=4$ -----

$c_{44}=0.9213177$

sdt/c_{44}

-----Varianza real del proceso $AR(1)$ -----

```
var1y=acf(X, lag.max = 0,type = c("covariance"), plot = FALSE, na.action = na.fail, demean = TRUE)
```

var1y

$\sqrt{1.09}$ ---colocar el valor de var1y----

----MÉTODO TRADICIONAL DESVIACIÓN ESTÁNDAR, $n=5$ y $\phi=0.2677$ -----

```
X<-arima.sim(n=100, list(ar=c(0.2677),ma=c(0)),sd=1)
datos=X
sample<-gl(20,5)
subgrupos <- qcc.groups(datos, sample)
sd=c()
x=c()
for(i in 1:20){
x[i]=sd(subgrupos[i,])
}
sd=c(sd, x)
sdt=mean(sd)
```

-----Desviación estándar muestral-----

sdt

-----Valores para constante c_4 , $n=5$ -----

$c_{45}=0.9399856$

sdt/c_{45}

-----Varianza real del proceso $AR(1)$ -----

```
var1y=acf(X, lag.max = 0,type = c("covariance"), plot = FALSE, na.action = na.fail, demean = TRUE)
```

```
var1y
sqrt(1.09)
```

-----MÉTODO TRADICIONAL RANGO, n=4 y phi=0.2677-----

```
X<-arima.sim(n=100, list(ar=c(0.2677),ma=c(0)),sd=1)
datos=X
sample<-gl(25,4)
subgrupos <- qcc.groups(datos, sample)
r=c()
x=c()
for(i in 1:25){
x[i]=diff(range(subgrupos[i,]))
}
r=c(r, x)
rt=mean(r)
```

-----Rango muestral-----

```
rt
```

-----Valores para constante d2, n=4 -----

```
d24=2.0587
```

```
rt/d24
```

-----Varianza real del proceso $AR(1)$ -----

```
var1y=acf(X, lag.max = 0,type = c("covariance"), plot = FALSE, na.action = na.fail, demean = TRUE)
```

```
var1y
```

```
sqrt(1.13)
```

-----MÉTODO TRADICIONAL RANGO, n=5 y phi=0.2677-----

```
X<-arima.sim(n=100, list(ar=c(0.2677),ma=c(0)),sd=1)
datos=X
sample<-gl(20,5)
subgrupos <- qcc.groups(datos, sample)
r=c()
x=c()
for(i in 1:20){
x[i]=diff(range(subgrupos[i,]))
}
r=c(r, x)
rt=mean(r)
```

-----Rango muestral-----

```
rt
```

-----Valores para constante d2, n=5 -----

```
d25=2.3259
```

```
rt/d25
```

-----Varianza real del proceso $AR(1)$ -----

```
var1y=acf(X, lag.max = 0,type = c("covariance"), plot = FALSE, na.action = na.fail, demean = TRUE)
```

```
var1y
```

```
sqrt(1.13)
```

Tabla 0.2 Constantes d_2 y d_3 para diferentes tamaños de muestra.

```
n=5;
n*(n-1)*NIntegrate[(r^2)*(CDF[NormalDistribution[0,1],y+r]-CDF[NormalDistribution[0,1],y])^(n-2)*PDF[NormalDistribution[0,1],y]*PDF[NormalDistribution[0,1],y+r]
,{r,0,Infinity},{y,Infinity,Infinity}]
```

Tabla 5.4 ARL para carta clásica \bar{X} , con diferentes valores de δ y n .

```
n = 1
arl =c()
k = seq(0,3, by=0.5)
for(i in 1:length(k) ){
p = 1- pnorm(-k[i]*sqrt(n) + 3) + pnorm(-k[i]*sqrt(n) - 3)
  a = 1/p
  arl = c(arl,a)
}
```

```
res1 = c(k,arl)
dim(res1) = c(length(k),2)
res1
```

```
n = 10
arl =c()
k = seq(0,3, by=0.5)
for(i in 1:length(k) ){
p = 1- pnorm(-k[i]*sqrt(n) + 3) + pnorm(-k[i]*sqrt(n) - 3)
  a = 1/p
  arl = c(arl,a)
}
```

```
res2 = c(k,arl)
dim(res2) = c(length(k),2)
res2
```

Tabla 0.3 ARL para carta de medias clásica, $n=4$.

*****CARTA DE MEDIAS para n=4*****

```
-----Variable: Altura total-----
datos=read.csv("datossanden.csv")
datos=datos$x1
sample<-gl(25,4)
subgrupos <- qcc.groups(datos, sample)
medias=c()
x=c()
for(i in 1:25){
x[i]=mean(subgrupos[i,])
}
```

```
medias=c(medias, x)
medias
medi=mean(medias)
medi
```



```
***35.4335
****Media muestral****
```

```
*****Estimación de la desviación estándar*****
datos=read.csv("datossanden.csv")
datos=datos$x1
sample<-gl(25,4)
subgrupos <- qcc.groups(datos, sample)
rangos=c()
rank=c()
for(i in 1:25){
    rank[i]=diff(range(subgrupos[i,]))
}
rangos=c(rangos, rank)
mean(rangos)
*** 0.5308
****Rango muestral****
****Calculando d2, n=4****
sigma=(mean(rangos))/2.059
sigma
****0.2577****
****Límites de control*****
n=4
ucl=medi+(3*sigma/sqrt(n))
lcl=medi-(3*sigma/sqrt(n))
medias=c()
for (i in 1:100)
{
    XT<-arima.sim(n=10000, list(ar=c(0.9),ma=c(0)),sd=sigma)+ (medi+0.5*sigma)
    mat=matrix(XT, nrow = 2500, ncol = 4, byrow = TRUE,dimnames = NULL)
    bolsa=c()
    x=c()
    for(i in 1:2500)
    {
        x[i]=mean(mat[i,])
        if (x[i]>=ucl||x[i]<=lcl) bolsa=c(bolsa,1) else bolsa=c(bolsa,0)
    }
    z=sum(bolsa)
    h=sort.list(bolsa, decreasing=TRUE)[1:z]
    h=c(0,h)
    me =mean(diff(h))
    medias=c(medias,me)
}
mean(medias)
```

Tabla 0.4 ARL para carta de medias clásica, n=5.

```
-----
*****CARTA DE MEDIAS para n=5*****
-----Variable: Altura total-----
datos=read.csv("datossanden.csv")
datos=datos$x1
```

```

sample<-gl(20,5)
subgrupos <- qcc.groups(datos, sample)
medias=c()
x=c()
for(i in 1:20){
x[i]=mean(subgrupos[i,])
}

medias=c(medias, x)
medias
medi=mean(medias)
medi
***35.4335
*****Media muestral*****

*****Estimación de la desviación estándar*****
datos=read.csv("datosanden.csv")
datos=datos$x1
sample<-gl(20,5)
subgrupos <- qcc.groups(datos, sample)
rangos=c()
rank=c()
for(i in 1:20){
rank[i]=diff(range(subgrupos[i,]))
}
rangos=c(rangos, rank)
mean(rangos)
*** 0.623
*****Rango muestral*****
*****Calculando d2, n=5*****
sigma=(mean(rangos))/2.3259
sigma
*****0.2678*****

*****Límites de control*****
n=5
ucl=medi+(3*sigma/sqrt(n))
lcl=medi-(3*sigma/sqrt(n))
medias=c()
for (i in 1:100)
{
XT<-arima.sim(n=10000, list(ar=c(0.9),ma=c(0)),sd=sigma)+ (medi+0.5*sigma)
mat=matrix(XT, nrow = 2000, ncol = 5, byrow = TRUE,dimnames = NULL)
bolsa=c()
x=c()
for(i in 1:2000)
{
x[i]=mean(mat[i,])
if (x[i]>=ucl||x[i]<=lcl) bolsa=c(bolsa,1) else bolsa=c(bolsa,0)
}
}
z=sum(bolsa)
h=sort.list(bolsa, decreasing=TRUE)[1:z]
h=c(0,h)

```

```

me =mean(diff(h))
medias=c(medias,me)
}
mean(medias)

```

Tabla 0.5 ARL para carta de rangos clásica.

#----- Calculando el ARL -----

```

n=5
fr <- function(z){( ( pnorm(z + 3*0.864 +2.326) - pnorm(z) )^(n-1) )*dnorm(z) }
d = integrate(fr, -Inf, Inf)
p =1-(n*d$value)
arl = 1/p
arl

```

*****Estimación de la desviación estándar*****

```

datos=read.csv("datossanden.csv")
datos=datos$x1
sample<-gl(20,5)
subgrupos <- qcc.groups(datos, sample)
rangos=c()
rank=c()
for(i in 1:20){
    rank[i]=diff(range(subgrupos[i,]))
}

```

```

rangos= rank
mean(rangos)

```

```

*** 0.623

```

*****Rango muestral*****

*****Calculando d2, n=5*****

```

sigma=(mean(rangos))/2.325929

```

```

sigma

```

```

****0.26785****

```

*****CARTA DE RANGOS, para rho=0.2, n=5*****

```

ucl=ran+(3*sigma*0.864)

```

```

lcl=0

```

```

medias=c()

```

```

bolsa=c()

```

```

r=c()

```

```

for (i in 1:100)

```

```

{

```

```

    Xt<-rnorm(10000, 0.623, sigma)

```

```

    x = filter(Xt, filter=c(0.2), method="recursive")

```

```

    mat=matrix(x, nrow = 2000, ncol = 5, byrow = TRUE,dimnames = NULL)

```

```

    for(i in 1:2000)

```

```

    {

```

```

        r[i]=diff(range(mat[i,]))

```

```

        if (r[i]>=ucl||r[i]<=lcl) bolsa=c(bolsa,1) else bolsa=c(bolsa,0)

```

```

    }

```

```

z=sum(bolsa)

```

```

h=sort.list(bolsa, decreasing=TRUE)[1:z]
h=c(0,h)
me =mean(diff(h))
medias=c(medias,me)
}
mean(medias)

```

Tabla 0.6 Valores críticos de la distribución ji-cuadrada.

-----Distribución ji-cuadrada para $\alpha=0.05$ -----

```

m=c()
z=c()
n=1:10
for (i in 1:10)
{
z[i]=qchisq(0.005, i, ncp=0, lower.tail = FALSE, log.p = FALSE)
}
m=matrix(c(n,z),10,2)
m

```

Tabla 0.7 ARL para carta de varianza clásica.

----- CARTA DE VARIANZA, n=2 rho=0-----

```

datos=read.csv("datossanden.csv")
datos=datos$x1
mean(datos)
sample<-gl(50,2)
subgrupos <- qcc.groups(datos, sample)
vari=c()
for(i in 1:50){
x[i]=var(subgrupos[i,])
}
vari=c(vari, x)
varianza=mean(vari)
varianza
0.090505

```

*****Límites de control*****

```

n=2
ucl=(varianza)*(7.879/(n-1))
bolsa=c()
medias=c()
for (i in 1:100)
{
Xt<-rnorm(10000,varianza,sqrt(varianza))
mat=matrix(Xt, nrow = 5000, ncol =2, byrow = TRUE,dimnames = NULL)
for(i in 1:5000)
{
xs[i]=var(mat[i,])
if (xs[i]>=ucl) bolsa=c(bolsa,1) else bolsa=c(bolsa,0)
}
}

```

```

z=sum(bolsa)
h=sort.list(bolsa, decreasing=TRUE)[1:z]
h=c(0,h)
me =mean(diff(h))
medias=c(medias,me)
}
mean(medias)

```

----- CARTA DE VARIANZA n=10, phi=0.3-----

```

datos=read.csv("datossanden.csv")
datos=datos$x1
mean(datos)
sample<-gl(10,10)
subgrupos <- qcc.groups(datos, sample)
vari=c()
for(i in 1:10){
x[i]=var(subgrupos[i,])
}
vari=c(vari, x)
varianza=mean(vari)
varianza

```

*****Límites de control*****

```

n=10
ucl=(varianza)*(23.589/(n-1))
bolsa=c()
medias=c()
for (i in 1:100)
{
    Xt<-rnorm(10000,varianza,sqrt(varianza))
    x = filter(Xt, filter=c(0.3), method="recursive")
    mat=matrix(x, nrow = 1000, ncol =10, byrow = TRUE,dimnames = NULL)
    for(i in 1:1000)
    {
        xs[i]=var(mat[i,])
        if (xs[i]>=ucl) bolsa=c(bolsa,1) else bolsa=c(bolsa,0)
    }
}
z=sum(bolsa)
h=sort.list(bolsa, decreasing=TRUE)[1:z]
h=c(0,h)
me =mean(diff(h))
medias=c(medias,me)
}
mean(medias)

```

Tabla 0.8 ARL para carta modificada de medias, n=4.

```

n=4
ucl=medi+(3*sigma/sqrt(n))
lcl=medi-(3*sigma/sqrt(n))
medias=c()

```

```

bolsa=c()
for (i in 1:100)
{
  for(i in c(1:2500) ){ dat = rnorm(4,medi,sigma )
  med = mean(dat)
  if ( med>=ucl||med<=lcl) bolsa=c(bolsa,1) else bolsa=c(bolsa,0) }
z=sum(bolsa)
h=sort.list(bolsa, decreasing=TRUE)[1:z]
h=c(0,h)
me =mean(diff(h))
medias=c(medias,me)
}
mean(medias)

```

Tabla 0.9 ARL para carta medias modificada, n=5.

```

n=5
ucl=medi+(3*sigma/sqrt(n))
lcl=medi-(3*sigma/sqrt(n))
medias=c()
bolsa=c()
for (i in 1:100)
{
  for(i in c(1:2000) ){ dat = rnorm(5,medi,sigma )
  med = mean(dat)
  if ( med>=ucl||med<=lcl) bolsa=c(bolsa,1) else bolsa=c(bolsa,0) }
z=sum(bolsa)
h=sort.list(bolsa, decreasing=TRUE)[1:z]
h=c(0,h)
me =mean(diff(h))
medias=c(medias,me)
}
mean(medias)

```

Tabla 0.10 Estimación de varianza para un proceso autocorrelacionado.

El valor esperado de la varianza muestral cuando el proceso es $AR(1)$ con parámetro ρ es:

$$E_{\rho}(S_i^2) = \gamma_0 \left\{ 1 - \frac{2}{n-1} \sum_{j=1}^n \left(1 - \frac{j}{n}\right) \rho^j \right\}$$

$$E_{\alpha_1}(S_i^2), n = 2$$

Instrucción para Wolfram Alpha.

1-((2/1)*((sum ((1-(j/2))*a^j), j=1 to 2))

$$E_{\rho}(S_i^2) = \gamma_0 \left\{ 1 - \frac{2}{n-1} \sum_{j=1}^2 \left(1 - \frac{j}{2}\right) \rho^j \right\} = \gamma_0 \{1 - \rho\}$$

$$E_{\rho}(S_i^2), n = 3$$

$$1 - ((2/2) * ((\sum_{j=1}^3 ((1 - (j/3)) * a^j))), j=1 \text{ to } 3))$$

$$E_{\rho}(S_i^2) = \gamma_0 \left\{ 1 - \frac{2}{3-1} \sum_{j=1}^3 \left(1 - \frac{j}{3} \right) \rho^j \right\} = \gamma_0 \left\{ -\frac{\rho^2}{3} - \frac{2\rho}{3} + 1 \right\}$$

$$E_{\rho}(S_i^2), n = 4$$

$$1 - ((2/3) * ((\sum_{j=1}^4 ((1 - (j/4)) * a^j))), j=1 \text{ to } 4))$$

$$0.083213 * (1 - (2/3)) * ((0.6^3/4) + ((0.6^2)/2) + ((3*0.6)/4))$$

$$E_{\rho}(S_i^2) = \gamma_0 \left\{ 1 - \frac{2}{4-1} \sum_{j=1}^4 \left(1 - \frac{j}{4} \right) \rho^j \right\} = \gamma_0 \left\{ 1 - \frac{2}{3} \left(\frac{\rho^3}{4} + \frac{\rho^2}{2} + \frac{3\rho}{4} \right) \right\}$$

$$E_{\rho}(S_i^2), n = 5$$

$$1 - ((2/4) * ((\sum_{j=1}^5 ((1 - (j/5)) * a^j))), j=1 \text{ to } 5))$$

$$1.1135 * ((0.5) * ((-0.1^4)/5) + (-2 * (0.1^3))/5) + (-3 * (0.1^2))/5 + ((-4 * (0.1))/5) + 1)$$

$$E_{\rho}(S_i^2) = \gamma_0 \left\{ 1 - \frac{2}{5-1} \sum_{j=1}^5 \left(1 - \frac{j}{5} \right) \rho^j \right\} = \gamma_0 \left\{ \frac{1}{2} \left(-\frac{\rho^4}{5} - \frac{2\rho^3}{5} - \frac{3\rho^2}{5} - \frac{4\rho}{5} \right) + 1 \right\}$$

$$E_{\rho}(S_i^2), n = 6$$

$$1 - ((2/5) * ((\sum_{j=1}^6 ((1 - (j/6)) * a^j))), j=1 \text{ to } 6))$$

$$E_{\rho}(S_i^2) = \gamma_0 \left\{ 1 - \frac{2}{6-1} \sum_{j=1}^6 \left(1 - \frac{j}{6} \right) \rho^j \right\} = \gamma_0 \left\{ 1 - \frac{2}{5} \left(\frac{\rho^5}{6} + \frac{\rho^4}{3} + \frac{\rho^3}{2} + \frac{2\rho^2}{3} + \frac{5\rho}{6} \right) \right\}$$

$$E_{\rho}(S_i^2), n = 7$$

$$1 - ((2/6) * ((\sum_{j=1}^7 ((1 - (j/7)) * a^j))), j=1 \text{ to } 7))$$

$$E_{\rho}(S_i^2) = \gamma_0 \left\{ 1 - \frac{2}{7-1} \sum_{j=1}^7 \left(1 - \frac{j}{7} \right) \rho^j \right\}$$

$$= \gamma_0 \left\{ \frac{1}{3} \left(-\frac{\rho^6}{7} - \frac{2\rho^5}{7} - \frac{3\rho^4}{7} - \frac{4\rho^3}{7} - \frac{5\rho^2}{7} - \frac{6\rho}{7} \right) + 1 \right\}$$

$$E_{\rho}(S_i^2), n = 8$$

$$1 - \left(\frac{2}{7} \right) \left(\sum_{j=1}^8 \left(1 - \frac{j}{8} \right) \rho^j \right)$$

$$E_{\rho}(S_i^2) = \gamma_0 \left\{ 1 - \frac{2}{8-1} \sum_{j=1}^8 \left(1 - \frac{j}{8} \right) \rho^j \right\}$$

$$= \gamma_0 \left\{ 1 - \frac{2}{7} \left(\frac{\rho^7}{8} + \frac{\rho^6}{4} + \frac{3\rho^5}{8} + \frac{\rho^4}{2} + \frac{5\rho^3}{8} + \frac{3\rho^2}{4} + \frac{7\rho}{8} \right) \right\}$$

$$E_{\rho}(S_i^2), n = 9$$

$$1 - \left(\frac{2}{8} \right) \left(\sum_{j=1}^9 \left(1 - \frac{j}{9} \right) \rho^j \right)$$

$$E_{\rho}(S_i^2) = \gamma_0 \left\{ 1 - \frac{2}{9-1} \sum_{j=1}^9 \left(1 - \frac{j}{9} \right) \rho^j \right\}$$

$$= \gamma_0 \left\{ \frac{1}{4} \left(-\frac{\rho^8}{9} - \frac{2\rho^7}{9} - \frac{\rho^6}{3} - \frac{4\rho^5}{9} - \frac{5\rho^4}{9} - \frac{2\rho^3}{3} - \frac{7\rho^2}{9} - \frac{8\rho}{9} \right) + 1 \right\}$$

$$E_{\rho}(S_i^2), n = 10$$

$$1 - \left(\frac{2}{9} \right) \left(\sum_{j=1}^{10} \left(1 - \frac{j}{10} \right) \rho^j \right)$$

$$E_{\rho}(S_i^2) = \gamma_0 \left\{ 1 - \frac{2}{10-1} \sum_{j=1}^{10} \left(1 - \frac{j}{10} \right) \rho^j \right\}$$

$$= \gamma_0 \left\{ 1 - \frac{2}{9} \left(\frac{\rho^9}{10} + \frac{\rho^8}{5} + \frac{3\rho^7}{10} + \frac{2\rho^6}{5} + \frac{\rho^5}{2} + \frac{3\rho^4}{5} + \frac{7\rho^3}{10} + \frac{4\rho^2}{5} + \frac{9\rho}{10} \right) \right\}$$

La varianza de la varianza muestral cuando el proceso es $AR(1)$ con parámetro ρ es:

$$Var_{\rho}(S_i^2) = \frac{2}{(n-1)^2} \gamma_0^2 \left\{ \sum_{i=1}^n \sum_{j=1}^n \rho^{2|i-j|} - \frac{2}{n} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \rho^{|i-j|+|i-k|} + \frac{1}{n^2} \left(\sum_{i=1}^n \sum_{j=1}^n \rho^{|i-j|} \right)^2 \right\}$$

$$Var_{\rho}(S_i^2), n = 2$$

Código en Wolfram Alpha.

```
(sum (sum a^(2*abs(i-j)), j=1 to 2), i=1 to 2)
sum (sum (sum a^(abs(i-j)+abs(i-k)), k=1 to 2), j=1 to 2), i=1 to 2)
((sum (sum a^(abs(i-j)), j=1 to 2), i=1 to 2))^2
(2)* 1.095098*(((2*(0.1^2))+2)+(- ((2*(0.1^2))+4*(0.1))+2))+((0.25)*(((2*0.1)+2)^2)))
```

$$Var_{\rho}(S_i^2) = \frac{2}{(2-1)^2} \gamma_0^2 \left\{ (2\rho^2 + 2) - \frac{2}{2} (2\rho^2 + 4\rho + 2) + \frac{1}{2^2} (2\rho + 2)^2 \right\}$$

$$Var_{\rho}(S_i^2), n = 3$$

```
(0.5)*0.854001*(((2*(0.1^4))+4*(0.1^2))+3)+(-
(2/3)*((2*(0.1^4))+4*(0.1^3))+10*(0.1^2))+8*(0.1))+3))+((1/9)*(((2*(0.1^2))+4*0.1)+3)^2)))
```

$$Var_{\rho}(S_i^2) = \frac{2}{(3-1)^2} \gamma_0^2 \left\{ (2\rho^4 + 4\rho^2 + 3) - \frac{2}{3} (2\rho^4 + 4\rho^3 + 10\rho^2 + 8\rho + 3) + \frac{1}{3^2} (2\rho^2 + 4\rho + 3)^2 \right\}$$

$$Var_{\rho}(S_i^2), n = 4$$

```
(2/9)* 0.7853291*(((2*(0.9^6))+4*(0.9^4))+6*(0.9^2))+4)-(2/4)*
((2*(0.9^6))+4*(0.9^5))+8*(0.9^4))+16*(0.9^3))+18*(0.9^2))+12*0.9)+4)+(1/(4^2))*
((2*(0.9^3))+4*(0.9^2))+6*0.9)+4)^2 )
```

$$Var_{\rho}(S_i^2) = \frac{2}{(4-1)^2} \gamma_0^2 \left\{ (2\rho^6 + 4\rho^4 + 6\rho^2 + 4) - \frac{2}{4} (2\rho^6 + 4\rho^5 + 8\rho^4 + 16\rho^3 + 18\rho^2 + 12\rho + 4) + \frac{1}{4^2} (2\rho^3 + 4\rho^2 + 6\rho + 4)^2 \right\}$$

$$Var_{\rho}(S_i^2), n = 5$$

$$\begin{aligned} & (2/16) * 1.004725 * ((2*(0.1^8)) + (4*(0.1^6)) + (6*(0.1^4)) + (8*(0.1^2)) + 5) - (2/5) * \\ & ((2*(0.1^8)) + (4*(0.1^7)) + (8*(0.1^6)) + (12*(0.1^5)) + (24*(0.1^4)) + \\ & (28*(0.1^3)) + (26*(0.1^2)) + (16*0.1) + 5) + (1/(5^2)) * ((2*(0.1^4)) + (4*(0.1^3)) + \\ & (6*(0.1^2)) + (8*0.1) + 5)^2 \end{aligned}$$

$$\begin{aligned} Var_{\rho}(S_i^2) = & \frac{2}{(5-1)^2} \gamma_0^2 \left\{ (2\rho^8 + 4\rho^6 + 6\rho^4 + 8\rho^2 + 5) \right. \\ & - \frac{2}{5} (2\rho^8 + 4\rho^7 + 8\rho^6 + 12\rho^5 + 24\rho^4 + 28\rho^3 + 26\rho^2 + 16\rho + 5) \\ & \left. + \frac{1}{5^2} (2\rho^4 + 4\rho^3 + 6\rho^2 + 8\rho + 5)^2 \right\} \end{aligned}$$

$$Var_{\rho}(S_i^2), n = 6$$

$$\begin{aligned} & (2/25) * 1.004725 * ((2*(0.1^{10})) + (4*(0.1^8)) + (6*(0.1^6)) + (8*(0.1^4)) + (10*(0.1^2)) + 6) - (2/6) * \\ & ((2*(0.1^{10})) + (4*(0.1^9)) + (8*(0.1^8)) + (12*(0.1^7)) + (18*(0.1^6)) + (32*(0.1^5)) + \\ & (40*(0.1^4)) + (40*(0.1^3)) + (34*(0.1^2)) + (20*0.1) + 6) + (1/(6^2)) * ((2*(0.1^5)) \\ & + (4*(0.1^4)) + (6*(0.1^3)) + (8*(0.1^2)) + (10*0.1) + 6)^2 \end{aligned}$$

$$\begin{aligned} Var_{\rho}(S_i^2) = & \frac{2}{(6-1)^2} \gamma_0^2 \left\{ (2\rho^{10} + 4\rho^8 + 6\rho^6 + 8\rho^4 + 10\rho^2 + 6) \right. \\ & - \frac{2}{6} (2\rho^{10} + 4\rho^9 + 8\rho^8 + 12\rho^7 + 18\rho^6 + 32\rho^5 + 40\rho^4 + 40\rho^3 + 34\rho^2 \\ & \left. + 20\rho + 6) + \frac{1}{6^2} (2\rho^5 + 4\rho^4 + 6\rho^3 + 8\rho^2 + 10\rho + 6)^2 \right\} \end{aligned}$$

$$Var_{\rho}(S_i^2), n = 7$$

$$\begin{aligned} & (2/36) * 1.004725 * ((2*(0.1^{12})) + (4*(0.1^{10})) + (6*(0.1^8)) + (8*(0.1^6)) + (10*(0.1^4)) + \\ & (12*(0.1^2)) + 7) - (2/7) * ((2*(0.1^{12})) + (4*(0.1^{11})) + (8*(0.1^{10})) \\ & + (12*(0.1^9)) + (18*(0.1^8)) + (24*(0.1^7)) + (42*(0.1^6)) + (52*(0.1^5)) + (56*(0.1^4)) + (52*(0.1^3)) + \\ & (42*(0.1^2)) + (24*0.1) + 7) + (1/(7^2)) * ((2*(0.1^6)) + (4*(0.1^5)) + (6*(0.1^4)) + (8*(0.1^3)) + \\ & (10*(0.1^2)) + (12*0.1) + 7)^2 \end{aligned}$$

$$\begin{aligned}
\text{Var}_\rho(S_i^2) = & \frac{2}{(7-1)^2} \gamma_0^2 \left\{ (2\rho^{12} + 4\rho^{10} + 6\rho^8 + 8\rho^6 + 10\rho^4 + 12\rho^2 + 7) \right. \\
& - \frac{2}{7} (2\rho^{12} + 4\rho^{11} + 8\rho^{10} + 12\rho^9 + 18\rho^8 + 24\rho^7 + 42\rho^6 + 52\rho^5 + 56\rho^4 \\
& + 52\rho^3 + 42\rho^2 + 24\rho + 7) \\
& \left. + \frac{1}{7^2} (2\rho^6 + 4\rho^5 + 6\rho^4 + 8\rho^3 + 10\rho^2 + 12\rho + 7)^2 \right\}
\end{aligned}$$

$$\text{Var}_\rho(S_i^2), n = 8$$

$$\begin{aligned}
& (2/49) * 1.004725 * (((2*(0.1^14)) + (4*(0.1^12)) + (6*(0.1^10)) \\
& + (8*(0.1^8)) + (10*(0.1^6)) + (12*(0.1^4)) + (14*(0.1^2)) + 8) - (2/8) * ((2*(0.1^14)) + (4*(0.1^13)) + \\
& (8*(0.1^12)) + (12*(0.1^11)) + (18*(0.1^10)) \\
& + (24*(0.1^9)) + (32*(0.1^8)) + (52*(0.1^7)) + (66*(0.1^6)) + (72*(0.1^5)) + \\
& (72*(0.1^4)) + (64*(0.1^3)) + (50*(0.1^2)) + (28*0.1) + 8) + (1/(8^2)) * ((2*(0.1^7)) + (4*(0.1^6)) \\
& + (6*(0.1^5)) + (8*(0.1^4)) + (10*(0.1^3)) + (12*(0.1^2)) + (14*0.1) + 8)^2)
\end{aligned}$$

$$\begin{aligned}
\text{Var}_\rho(S_i^2) = & \frac{2}{(8-1)^2} \gamma_0^2 \left\{ (2\rho^{14} + 4\rho^{12} + 6\rho^{10} + 8\rho^8 + 10\rho^6 + 12\rho^4 + 14\rho^2 + 8) \right. \\
& - \frac{2}{8} (2\rho^{14} + 4\rho^{13} + 8\rho^{12} + 12\rho^{11} + 18\rho^{10} + 24\rho^9 + 32\rho^8 + 52\rho^7 \\
& + 66\rho^6 + 72\rho^5 + 72\rho^4 + 64\rho^3 + 50\rho^2 + 28\rho + 8) \\
& \left. + \frac{1}{8^2} (2\rho^7 + 4\rho^6 + 6\rho^5 + 8\rho^4 + 10\rho^3 + 12\rho^2 + 14\rho + 8)^2 \right\}
\end{aligned}$$

$$\text{Var}_\rho(S_i^2), n = 9$$

$$\begin{aligned}
& (2/64) * 1.004725 * (((2*(0.1^16)) + (4*(0.1^14)) + (6*(0.1^12)) + (8*(0.1^10)) \\
& + (10*(0.1^8)) + (12*(0.1^6)) + (14*(0.1^4)) + (16*(0.1^2)) + 9) - (2/9) * ((2*(0.1^16)) + (4*(0.1^15)) + \\
& (8*(0.1^14)) + (12*(0.1^13)) + (18*(0.1^12)) + (24*(0.1^11)) + (32*(0.1^10)) \\
& + (40*(0.1^9)) + (64*(0.1^8)) + (80*(0.1^7)) + (90*(0.1^6)) + (92*(0.1^5)) + \\
& (88*(0.1^4)) + (76*(0.1^3)) + (58*(0.1^2)) + (32*0.1) + 9) + (1/(9^2)) * ((2*(0.1^8)) + (4*(0.1^7)) + \\
& (6*(0.1^6)) + (8*(0.1^5)) + (10*(0.1^4)) + (12*(0.1^3)) + (14*(0.1^2)) + (16*0.1) + 9)^2)
\end{aligned}$$

$$\begin{aligned}
Var_{\rho}(S_i^2) = & \frac{2}{(9-1)^2} \gamma_0^2 \left\{ (2\rho^{16} + 4\rho^{14} + 6\rho^{12} + 8\rho^{10} + 10\rho^8 + 12\rho^6 + 14\rho^4 + 16\rho^2 \right. \\
& + 9) \\
& - \frac{2}{9} (2\rho^{16} + 4\rho^{15} + 8\rho^{14} + 12\rho^{13} + 18\rho^{12} + 24\rho^{11} + 32\rho^{10} + 40\rho^9 \\
& + 64\rho^8 + 80\rho^7 + 90\rho^6 + 92\rho^5 + 88\rho^4 + 76\rho^3 + 58\rho^2 + 32\rho + 9) \\
& \left. + \frac{1}{9^2} (2\rho^8 + 4\rho^7 + 6\rho^6 + 8\rho^5 + 10\rho^4 + 12\rho^3 + 14\rho^2 + 16\rho + 9)^2 \right\}
\end{aligned}$$

$$Var_{\rho}(S_i^2), n = 10$$

$$\begin{aligned}
& (2/81)*1.004725*(((2*(0.1^18))+4*(0.1^16))+6*(0.1^14))+8*(0.1^12))+10*(0.1^10)) \\
& +(12*(0.1^8))+14*(0.1^6))+16*(0.1^4))+18*(0.1^2))+10)-(2/10)*((2*(0.1^18))+4*(0.1^17))+ \\
& (8*(0.1^16))+12*(0.1^15))+18*(0.1^14))+24*(0.1^13))+32*(0.1^12))+40*(0.1^11))+ \\
& (50*(0.1^10))+76*(0.1^9))+96*(0.1^8))+108*(0.1^7))+114*(0.1^6))+112*(0.1^5))+ \\
& (104*(0.1^4))+88*(0.1^3))+66*(0.1^2))+36*0.1)+10)+(1/(10^2))*((2*(0.1^9))+4*(0.1^8)) \\
& +(6*(0.1^7))+8*(0.1^6))+10*(0.1^5))+12*(0.1^4))+14*(0.1^3))+16*(0.1^2))+18*0.1)+10)^2)
\end{aligned}$$

$$\begin{aligned}
Var_{\rho}(S_i^2) = & \frac{2}{(10-1)^2} \gamma_0^2 \left\{ (2\rho^{18} + 4\rho^{16} + 6\rho^{14} + 8\rho^{12} + 10\rho^{10} + 12\rho^8 + 14\rho^6 + 16\rho^4 \right. \\
& + 18\rho^2 + 10) \\
& - \frac{2}{10} (2\rho^{18} + 4\rho^{17} + 8\rho^{16} + 12\rho^{15} + 18\rho^{14} + 24\rho^{13} + 32\rho^{12} + 40\rho^{11} \\
& + 50\rho^{10} + 76\rho^9 + 96\rho^8 + 108\rho^7 + 114\rho^6 + 112\rho^5 + 104\rho^4 + 88\rho^3 \\
& + 66\rho^2 + 36\rho + 10) \\
& \left. + \frac{1}{10^2} (2\rho^9 + 4\rho^8 + 6\rho^7 + 8\rho^6 + 10\rho^5 + 12\rho^4 + 14\rho^3 + 16\rho^2 + 18\rho \right. \\
& \left. + 10)^2 \right\}
\end{aligned}$$

Tabla 0.11 ARL para carta modificada de varianzas.

```
set.seed(150)
datos=(arima.sim(n=100, list(ar=c(0.1)), sd=sqrt(1)))+35.4335
sample<-gl(10,10)
subgrupos <- qcc.groups(datos, sample)
vari=c()
for(i in 1:10){
v[i]=var(subgrupos[i,])
}
vari=c(vari, v)
vari
varianza=mean(vari)
varianza
```

-----Límites de control-----

```
n=10
ucl=(0.2226)*(23.589/(n-1))
medias=c()
for (i in 1:100)
{
bolsa=c()

      for (i in 1:1000)
      {
Xt=(rnorm(10, 0.2226,sqrt(0.2226)))+35.4335
xs=var(Xt)
if (xs>=ucl) bolsa=c(bolsa,1) else bolsa=c(bolsa,0)
}
z=sum(bolsa)
h=sort.list(bolsa, decreasing=TRUE)[1:z]
h=c(0,h)
me =mean(diff(h))
medias=c(medias,me)
}
mean(medias)
```

APÉNDICE

Distribución de un estadístico de orden.

La expresión X_1, X_2, \dots, X_n es una muestra aleatoria simple con función de densidad de probabilidad (fdp), $f(x)$ y función de distribución acumulada (fda), $F(x)$. Si se observa una realización concreta de tal muestra se escribe (x_1, x_2, \dots, x_n) . De manera que, $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$ son los estadísticos de orden, en donde $x_{(1)} = \min\{x_1, \dots, x_n\}$, $x_{(2)}$ es el segundo valor más pequeño de la realización y así sucesivamente hasta tener el máximo $x_{(n)} = \max\{x_1, \dots, x_n\}$. El rango se obtiene mediante $X_{(n)} - X_{(1)}$.

En una carta de control para rangos se tiene el rango promedio, \bar{R} , de varias muestras iniciales. Se sabe que, para muestras de tamaño n : $E(R_i/d_2) = \sigma$, donde $d_2 = \int_{-\infty}^{\infty} [1 - \phi^n(Y) - (1 - \phi(Y))^n] * dy$. De acuerdo con *Balakrishnan y Nagaraja (2008)*, Y_j es el j -ésimo estadístico de orden de una muestra aleatoria $\{X_1, \dots, X_n\} \sim F_X()$ y la siguiente expresión corresponde a la función de densidad conjunta de los n estadísticos de orden:

$$f_{X_j}(X_j) = \frac{n!}{(j-1)!(n-j)!} [F(X_j)]^{j-1} [1 - F(X_j)]^{n-j} * f(X_j) \quad a < X_j < b \quad (1)$$

donde (a, b) son el soporte de la función de densidad de $\{X_1, \dots, X_n\}$. Entonces la *función de densidad* cuando $j = 1$ y $j = n$ son:

$$\begin{aligned} f_{X_1}(X) &= \frac{n!}{(n-1)!} * [F(X)]^0 * [1 - F(X)]^{n-1} * f(X) \\ &= n * [1 - F(X)]^{n-1} * f(X) \quad \text{para } a < X < b \end{aligned} \quad (2)$$

$$y f_{X_n}(X) = n * [F(X)]^{n-1} * f(X) \quad \text{para } a < X < b \quad (3)$$