

**EL EFECTO DE LA ESTIMACIÓN DE LA  
MATRIZ DE COVARIANZAS EN EL  
DESEMPEÑO DE LA CARTA  $|S|$**

**Maryoris Mairén Pérez Sánchez**



**UNIVERSIDAD DE CÓRDOBA  
FACULTAD DE CIENCIAS BÁSICAS  
DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA  
MONTERÍA - CÓRDOBA  
2020**

# EL EFECTO DE LA ESTIMACIÓN DE LA MATRIZ DE COVARIANZAS EN EL DESEMPEÑO DE LA CARTA $|S|$

Maryoris Mairén Pérez Sánchez

Trabajo presentado como requisito parcial para optar al título de  
**ESTADÍSTICO**

Director:

**Víctor Hugo Morales Ospina**



UNIVERSIDAD DE CÓRDOBA  
FACULTAD DE CIENCIAS BÁSICAS  
DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA  
MONTERÍA - CÓRDOBA  
2020

UNIVERSIDAD DE CÓRDOBA  
FACULTAD DE CIENCIAS BÁSICAS  
DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA  
PROGRAMA DE ESTADÍSTICA

Los jurados abajo firmantes certifican que han leído y dan calificación al trabajo de grado titulado EL EFECTO DE LA ESTIMACIÓN DE LA MATRIZ DE COVARIANZA EN EL DESEMPEÑO DE LA CARTA  $|S|$  presentado por MARYORIS MAIREN PÉREZ SÁNCHEZ como requisito parcial para optar el título de ESTADÍSTICO.

Fecha: Julio de 2020

Director: \_\_\_\_\_  
Víctor Morales Ospina

Jurados: \_\_\_\_\_  
Hugo Brango García

\_\_\_\_\_  
Manuel Treco Hernández



*A mis padres*

*Roberto Pérez Arias & Milena Sánchez Coronel*

# Índice general

<b>Resumen</b>	<b>VIII</b>
<b>Abstract</b>	<b>IX</b>
<b>Agradecimientos</b>	<b>I</b>
<b>Introducción</b>	<b>1</b>
<b>1. Marco teórico</b>	<b>4</b>
1.1. Conceptos de calidad . . . . .	4
1.2. Control estadístico de procesos . . . . .	9
1.2.1. Histograma . . . . .	10
1.2.1.1. Ventajas . . . . .	12
1.2.1.2. Desventajas . . . . .	13
1.2.2. Hoja de Verificación . . . . .	13
1.2.2.1. Ventajas . . . . .	14
1.2.2.2. Desventajas . . . . .	15
1.2.3. Diagrama de pareto . . . . .	15
1.2.3.1. Ventajas . . . . .	16
1.2.3.2. Desventajas . . . . .	16
1.2.4. Diagrama Causa - Efecto . . . . .	17
1.2.4.1. Ventajas . . . . .	18

1.2.4.2.	Desventajas . . . . .	18
1.2.5.	Diagrama de concentración de defectos . . . . .	19
1.2.5.1.	¿Dónde ocurren los defectos? . . . . .	19
1.2.6.	Diagrama de dispersión . . . . .	20
1.2.7.	Diagrama de control . . . . .	22
1.3.	Carta de control . . . . .	22
1.3.1.	Definición . . . . .	22
1.3.2.	Tipos de variabilidad . . . . .	24
1.3.2.1.	Variabilidad debida a causas no asignables (comunes o aleatorias) . . . . .	24
1.3.2.2.	Variabilidad debida a causas asignables (especiales o no aleatorias) . . . . .	25
1.3.3.	Construcción de una carta de control . . . . .	26
1.3.4.	Criterios para concluir cuando un proceso está fuera de control . . .	28
1.3.4.1.	¿Cuándo un proceso está bajo control? . . . . .	28
1.3.4.2.	¿Cuándo un proceso está fuera de control? . . . . .	29
1.3.5.	Probabilidad de falsa alarma ( $\alpha$ ) . . . . .	31
1.3.6.	Longitud de corrida (RL) . . . . .	31
1.3.6.1.	ARL . . . . .	32
1.3.6.2.	$ARL_0$ . . . . .	32
1.3.6.3.	$ARL_1$ . . . . .	32
1.3.7.	Tipos de cartas de control . . . . .	33
1.3.7.1.	Carta de control para variables . . . . .	33
1.3.7.2.	Carta de control para atributos . . . . .	33
1.3.8.	Estimación de los parámetros en una carta de control . . . . .	34
1.3.9.	Fases de una carta de control . . . . .	35
1.3.9.1.	Fase I . . . . .	35
1.3.9.2.	Fase II . . . . .	35

<b>2. Estado del arte</b>	<b>36</b>
<b>3. Estudio de simulación y resultados</b>	<b>42</b>
3.1. Escenario 1 . . . . .	42
3.1.1. Resultados . . . . .	42
3.1.2. Análisis gráfico . . . . .	45
3.2. Escenario 2 . . . . .	46
3.2.1. Resultados . . . . .	46
3.2.2. Análisis gráfico . . . . .	47
<b>4. Aplicación</b>	<b>52</b>
4.1. Análisis exploratorio de los datos . . . . .	53
4.2. Cartas de control univariadas . . . . .	56
4.3. Carta de control $T^2$ Hotelling . . . . .	57
4.4. Carta de control $ S $ . . . . .	58
<b>5. Conclusiones y recomendaciones</b>	<b>61</b>
5.1. Conclusiones . . . . .	61
5.2. Recomendaciones para futuras investigaciones . . . . .	62
<b>Bibliografía</b> . . . . .	<b>63</b>
<b>A. Demostraciones</b>	<b>68</b>
<b>B. Constantes para Gráficos de Control</b>	<b>70</b>
<b>C. Código de simulación en R</b>	<b>71</b>

# Lista de Tablas

- 1.1. Parámetros más comunes a controlar y sus estimadores. . . . . 28
- 1.2. Clasificación de las cartas de control. . . . . 33
  
- 3.1. Proporción de falsa alarma para la carta  $|S|$  cuando  $\Sigma$  es estimada, datos simulados bajo el Escenario 1. . . . . 44
- 3.2. Proporción de falsa alarma para la carta  $|S|$  cuando  $\Sigma$  se estima, datos simulados bajo el Escenario 2. . . . . 46
  
- 4.1. Análisis descriptivo de concentraciones de  $PM_{2.5}$  y  $PM_{10}$ . . . . . 54
- 4.2. Proporción de falsa alarma para la carta  $|S|$ . . . . . 60
  
- B.1. Constante para gráfico de control. . . . . 70

# Lista de Figuras

- 1.1. Ciclo PDCA - Ciclo Deming. . . . . 8
- 1.2. Trilogía de Juran. . . . . 8
- 1.3. Catorce pasos de mejoramiento de calidad de Crosby. . . . . 9
- 1.4. Hoja de verificación. . . . . 14
- 1.5. Ley de Pareto. . . . . 15
- 1.6. Diagrama de Causa - Efecto. . . . . 17
- 1.7. Áreas con defectos en determinado tiempo. . . . . 20
- 1.8. Diagrama de dispersión. . . . . 20
- 1.9. Tipos de correlación. . . . . 21
- 1.10. Estructura de un gráfico de control. . . . . 24
- 1.11. Probabilidad y causas de variación asignables. . . . . 25
- 1.12. Carta de control genérica. . . . . 27
- 1.13. Reglas de Western Electric. . . . . 30
  
- 3.1. Resultados bajo el resultado 1, cuando el número de muestras ( $m$ ) es fijo  
y el tamaño de los subgrupos ( $n$ ) varían. . . . . 45
- 3.2. Resultados bajo el escenario 1, cuando el tamaño de los subgrupos ( $n$ ) es  
fijo y el número de muestras ( $m$ ) varían. . . . . 45
- 3.3. Resultados bajo el escenario 2, cuando el número de muestras ( $m$ ) es fijo  
y el tamaño de los subgrupos ( $n$ ) varían. . . . . 47
- 3.4. Resultados bajo el escenario 2, cuando el tamaño de los subgrupos ( $n$ ) es  
fijo y el número de muestras ( $m$ ) varían. . . . . 48

3.5. Diagrama de dispersión de los datos simulados. . . . .	51
4.2. Diagrama de dispersión, diagrama de caja e histograma para $PM_{10}$ con los datos de PM2008MS8. . . . .	55
4.1. Diagrama de dispersión, diagrama de caja e histograma para $PM_{2.5}$ con los datos de PM2008MS8. . . . .	55
4.3. Matriz de gráfico de dispersión entre las concentraciones de $PM_{2.5}$ y $PM_{10}$ con los datos de PM2008MS8. . . . .	56
4.4. Cartas de Control S para $PM_{2.5}$ y $PM_{10}$ . . . . .	57
4.5. Carta de Control $T^2$ de Hotelling para las concentraciones de $PM_{2.5}$ y $PM_{10}$ en la estación MS8. . . . .	58
4.6. Carta da control multivariada $ S $ para $MP_{2.5}$ y $MP_{10}$ . . . . .	59
4.7. Distribución bootstrap para los resultados. . . . .	60

# Resumen

Cuando se diseña una carta de control, lo usual es que los parámetros deban ser estimados a partir de una muestra preliminar de observaciones. Numerosos estudios han evidenciado el efecto que la estimación de los parámetros puede tener sobre el desempeño de las cartas de control. Se ha encontrado que sólo para muestras preliminares relativamente grandes, el desempeño de las cartas de control con parámetros estimados puede considerarse satisfactorio, en términos de la tasa de falsas alarmas que la carta produce en Fase II. En este trabajo evaluamos el efecto de la estimación de parámetros en el desempeño de la carta  $|S|$ . Para esto, se determinó de manera aproximada el número de muestras preliminares y el tamaño de subgrupos a partir de los cuales el desempeño de la carta se considera adecuado; también se ajustó un modelo que permite relacionar el tamaño de la muestra preliminar y de los subgrupos, con la tasa de falsas alarmas observada. Para nuestro estudio, se usaron procedimientos de simulación que permitieron observar la tasa de falsas alarmas para distintos tamaños de la muestra preliminar.

**PALABRAS CLAVES:** Carta  $|S|$ , Tamaño de muestra, subgrupos, estimación de parámetros.

# Abstract

When designing a control chart, it is usual that the parameters should be estimated from a preliminary sample of observations. Numerous studies have shown the effect that the estimation of the parameters can have on the performance of the control charts. It has been found that only for relatively large preliminary samples, the performance of control charts with estimated parameters can be considered satisfactory, in terms of the rate of false alarms that the chart produces in Phase II. In this work we evaluate the effect of the estimation of parameters on the performance of the  $|S|$  chart. For this, the number of preliminary samples and the size of subgroups from which the performance of the chart is considered adequate were determined approximately, finally a model is was adjusted that allows the size of the preliminary sample to be related to the false alarm rate observed. For our study, simulation procedures were used to observe the false alarm rate for different sizes of the preliminary sample.

**keywords:**  $|S|$  chart, sample size, subgroups, parameter estimation.

# Agradecimientos

Principalmente a **DIOS**, por darme la vida y cada una de las bendiciones en ella.

A mi director, Víctor Morales Ospina, por su paciencia, comprensión, esmero, constancia, dedicación, motivación y disponibilidad para que todo este proceso se diera de la mejor manera y con los mejores resultados.

A los profesores del Programa de Estadísticas de la Universidad de Córdoba por ser un ejemplo a seguir tanto en lo profesional como en lo personal, muy especialmente al profesor Mario Morales Rivera, Osnamir Bru Cordero, Juana Robles Gonzales, Kevin Pérez, Roger Tovar y Guillermo Martínez.

A mis amigos y compañeros de estudio del programa de Estadística, por lo compartido en los últimos años, en especial a Antonio Zumaque, Edgar Ramos, Eduardo Moya, Einer Mejía, Javier Blanco, Lili Bautista, Katy Pacheco, Manuel Hernández, y María Martínez.

# Introducción

En la construcción de una carta de control, uno de los aspectos fundamentales es la elección de ciertas características que determinan el futuro desempeño de ésta. Una de estas características es la proporción de falsas alarmas ( $\alpha$ ) asociada a la carta. Para algunos esquemas de monitoreo, esta característica está plenamente determinada cuando los parámetros de la carta son conocidos. Por ejemplo, es bien conocido que la tasa de falsas alarmas asociada con la carta tradicional  $\bar{x}$  para monitorear la media, es igual a 0.0027 cuando se usan límites  $3\sigma$  y se conocen los parámetros. En la práctica, sin embargo, lo usual es que estos parámetros sean estimados a partir de una muestra preliminar de observaciones. Numerosos estudios tales como [Rivera \(2011\)](#) y [Blanco \(2012\)](#) han evidenciado el efecto que la estimación de parámetros tiene sobre el desempeño de las cartas de control. Se ha encontrado que sólo para muestras preliminares relativamente grandes, el desempeño de las cartas con parámetros estimados es similar al que tendrían si los parámetros fueran conocidos. Si se utilizan tamaños de muestra preliminar que no garanticen un adecuado funcionamiento de las cartas de control, se tendrían procesos monitoreados de manera inadecuada, lo que llevaría a la toma de decisiones equivocadas.

Cuando se usan cartas de control en procedimientos de monitoreo se pueden distinguir claramente dos Fases: Fase I y Fase II. De acuerdo con [Mahmoud y Woodall \(2004\)](#), el principal interés en Fase I, es el análisis de un conjunto histórico de datos del proceso para conocer la variación y determinar la estabilidad de este. A su vez, el principal objetivo del análisis en Fase II, es detectar rápidamente cambios en el proceso a partir de los parámetros en control estimados en Fase I. De acuerdo con estos autores, en Fase I las cartas de control se comparan de acuerdo con la probabilidad de decidir si el proceso es estable o no. [Zhao y Driscoll \(2016\)](#) señalan que para ciertos casos, el número de muestras necesarias en Fase I pueden llegar a 10.000. Además, mencionan que cuando se diseña una

carta de control, los parámetros frecuentemente no se conocen, y deben ser estimados a priori, antes de hacer el monitoreo del proceso. Esta estimación ocurre en la Fase I del proceso. En esta Fase, se hace un análisis retrospectivo de los datos del proceso, donde los parámetros deben ser estimados para establecer los límites de control que serán usados en el futuro monitoreo del proceso. Autores tales como [Chakraborti y Human \(2006\)](#); [Raubenheimer y Van der Merwe \(2015\)](#) estudiaron el efecto de la estimación de parámetros en el desempeño de las cartas de control  $p$  y  $c$ . [Testik \(2007\)](#) explora el efecto del error de estimación en una carta Poisson CUSUM y ofrece algunas recomendaciones acerca de los tamaños de muestra. [Jensen et al. \(2006\)](#) y [Psarakis et al. \(2014\)](#) ofrecen revisiones de literatura sobre el desempeño de las cartas de control con parámetros estimados.

[Chen et al. \(2016\)](#) estudian el efecto que las estimaciones en Fase I pueden tener en el desempeño de las cartas en Fase II, en el contexto de monitoreo de perfiles. En el área de las cartas multivariadas existen pocos estudios al respecto, uno de ellos fue desarrollado por [Aly et al. \(2016\)](#) quienes estudiaron el rendimiento de las cartas de control EWMA multivariadas (MAEWMA) con parámetros estimados para monitorear el vector de medias.

En los procesos multivariados, de forma análoga a los del caso univariado, frecuentemente resulta de interés monitorear la variabilidad. La carta de control de la varianza generalizada permite monitorear la variabilidad de procesos multivariados y sus límites de control están definidos mediante las expresiones:

$$\begin{aligned} LCI &= |\Sigma| \left( b_1 - 3\sqrt{b_2} \right), \\ LC &= b_1 |\Sigma|, \\ LCS &= |\Sigma| \left( b_1 + 3\sqrt{b_2} \right), \end{aligned} \tag{1}$$

donde,  $b_1$  y  $b_2$  con constantes definidas por:

$$\begin{aligned} b_1 &= \frac{1}{(n-1)^p} \prod_{i=1}^p (n-i), \\ b_2 &= \frac{1}{(n-1)^{2p}} \prod_{i=1}^p (n-i) \left[ \prod_{j=1}^p (n-j+2) - \prod_{j=1}^p (n-j) \right], \end{aligned} \tag{2}$$

con  $n$ , el tamaño de muestra en cada subgrupo y  $p$  el número de variables que se tienen en cuenta en el proceso.

Sin embargo, en la práctica, es poco usual que  $\Sigma$  sea conocida por lo que se requiere estimarla, es usual tomar  $\bar{S}$  como estimador, donde  $\bar{S}$  se define como el promedio de los  $S_i$  obtenidos en cada punto de muestreo. Los límites de control de la carta de la varianza generalizada cuando la matriz de covarianzas asociada al proceso no es conocida, están dados por:

$$\begin{aligned} LCI &= |S| \left( 1 - 3 \frac{\sqrt{b_2}}{b_1} \right), \\ LC &= |S|, \\ LCS &= |S| \left( 1 + 3 \frac{\sqrt{b_2}}{b_1} \right). \end{aligned} \tag{3}$$

De acuerdo con [Montgomery y Wadsworth \(1972\)](#) la mayor parte de la distribución esta contenida en el intervalo  $\mathbb{E}(|S|) \pm \sqrt{\text{Var}(|S|)}$ . Sin embargo, el cumplimiento de esta propiedad esta sujeto a la estimación que se haga de  $\Sigma$ . Una pobre estimación de  $\Sigma$  no permitirá el cumplimiento de esta propiedad, y derivará en un pobre esquema de monitoreo.

El objetivo principal de este trabajo es evaluar el efecto de la estimación de la matriz de covarianzas en el desempeño de la carta  $|S|$ , como función del tamaño de la muestra preliminar ( $m$ ) y del número de muestras ( $n$ ) utilizados en la estimación.

La estructura del presente trabajo es la siguiente: En el capítulo 1, se hace una revisión de los conceptos básicos relacionados con el concepto de calidad y el control estadístico de un proceso. En el capítulo 2, se hace una revisión de literatura sobre los resultados fundamentales que se han obtenido con relación al problema en estudio. En el capítulo 3, se presentan los diferentes escenarios bajos los cuales se realizará el estudio de simulación y se hace una breve descripción del proceso a ejecutar. En el capítulo 4, se muestra una aplicación de la carta para varianza generalizada  $|S|$  con datos reales. En el capítulo 5, se presentan las conclusiones y recomendaciones del trabajo. Además, se ofrecen algunas ideas para futuros trabajos.

# Capítulo 1

## Marco teórico

En esta sección se hace una revisión de los conceptos teóricos relacionados con el monitoreo de procesos, principalmente los relacionados con las cartas de control, se dan definiciones básicas y se muestran algunos desarrollos.

### 1.1. Conceptos de calidad

El concepto de Calidad<sup>1</sup> es un concepto cambiante que ha venido evolucionando a través del tiempo, por eso conviene conocer un poco de ello y saber las distintas etapas por las que ha ido pasando, así como los conceptos más importantes. Se puede hablar de seis etapas muy claras en la evolución del concepto:

1. Etapa artesanal: donde la calidad suponía hacer las cosas bien a cualquier costo.

---

<sup>1</sup>En una feria de pueblo un vendedor encaramado en su tarima, pregonaba ante un corrillo de parroquianos, las bondades de sus mercancías. Mientras exhibía unas camisas, decía: “Lleve esta finísima prenda por unos pocos pesos, mire su diseño, su tela aguanta el uso y el abuso por cinco años, y como si fuera poco, sus botones de alta calidad están garantizados por veinticinco años o más. Llévela, no por cinco mil pesos, ni por cuatro mil . . . ”

Un parroquiano del corrillo le manifestó a su vecino: “Esa prenda sería de mejor calidad si sus botones fueran de más mala calidad”.

Rendón (2013)

2. Etapa de la industrialización: donde el concepto de calidad fue sustituido por el de producción, hacer muchas cosas y muy deprisa sin importar con que calidad.
3. Etapa de control final: en esta etapa lo importante ya no era la cantidad de producto fabricado, sino que el cliente lo recibiera según sus especificaciones.
4. Etapa de control en proceso: Los defectuosos que se encontraban durante la inspección final, no sólo se producían durante el proceso de fabricación, sino que también eran provocados por el mal estado o la no conformidad de la materia prima utilizada.
5. Etapa de control en diseño: se tenía controlado el proceso y se adoptaban acciones correctoras y preventivas, pero se seguían detectando problemas de calidad que aparecían durante la vida útil del producto y que no eran imputables ni a la materia prima, ni a las máquinas, ni a la mano de obra, ni al proceso.
6. Mejora continua: en el mercado actual para ser competitivos, hay que dirigirse hacia la excelencia y eso sólo se consigue a través de la mejora continua de los productos y/o servicios.

El concepto más antiguo de calidad del que se tiene conocimiento es en el código de [Hammurabi \(1970\)](#), ley N° 229 (primer conjunto de leyes de la historia, se enumeran las leyes que ha recibido del dios Marduk para fomentar el bienestar entre las gentes) que cita lo siguiente:

“Si un albañil construye una casa para un hombre, y su trabajo no es fuerte y la casa se derrumba matando a su dueño, el albañil será condenado a muerte”.

Se puede decir, que se refería a algo meramente funcional, un producto o servicio se consideraba de calidad, con el simple hecho de cumplir con su propósito.

[Shewhart \(1924\)](#), físico, ingeniero y estadístico estadounidense, conocido como el padre del control estadístico de la calidad, entendía la calidad como un problema de variación, el cual puede ser controlado y prevenido mediante la eliminación a tiempo de las causas que lo provocan (cartas de control). En específico, estudiaba el porqué de las variaciones en una producción para corregirlas antes que estas se repitieran. [Feigenbaum \(1991\)](#), empresario estadounidense y experto en control de calidad declaró que la calidad es una determinación

del cliente, no una determinación del ingeniero, ni de mercadeo, ni del gerente general. Está basada en la experiencia actual del cliente con los productos o servicios. Por lo tanto, la calidad de un producto era determinada por la satisfacción del cliente en cuanto al producto final, propone los siguientes principios fundamentales acerca de la calidad:

- Definir políticas y objetivos de calidad.
- Orientarse al cliente.
- Integrar todas las actividades de la compañía.
- Aclarar el papel de cada miembro de la empresa.
- Establecer un sistema de aseguramiento de proveedores.
- Identificar por completo el equipo de calidad.
- Definir un flujo efectivo de información.
- Establecer un sistema de costos de calidad.
- Ser efectivo en las acciones correctivas.
- Mantener un control continuo del sistema.

[Deming \(1982\)](#), estadístico estadounidense, profesor universitario, autor de textos, consultor y difusor del concepto de calidad total, asegura que la calidad es el grado predecible de uniformidad y fiabilidad a un bajo costo y que se ajuste a las necesidades del mercado. Su estrategia se basaba en catorce puntos claves:

- Crear un propósito de mejora del producto y del servicio.
- Adoptar una filosofía que elimine los niveles aceptados de errores.
- Suspender la dependencia de la inspección mediante el control del proceso.
- Terminar con la práctica de hacer negocio sobre la base del precio.
- Mejorar constantemente y para siempre el sistema de trabajo mediante la planificación y mediante la aplicación de una metodología concreta en la producción.

- Aplicar métodos de entrenamiento en el trabajo.
- Adoptar un concepto de liderazgo que permita fomentar el trabajo en equipo y lograr la mayor productividad.
- Eliminar el miedo y fomentar la confianza para que todos trabajen con efectividad.
- Romper las barreras entre los departamentos de la empresa. Todos forman parte del mismo equipo y deben trabajar en la misma dirección.
- Eliminar eslóganes y metas encaminadas a incrementar la productividad si no se proporcionan los medios adecuados para ellos.
- Eliminar cuotas numéricas. Si la principal meta es la cantidad, la calidad se ve afectada.
- Eliminar las barreras que se encuentran entre el trabajador y el derecho a sentirse orgulloso de su trabajo.
- Instituir programas de formación para mejorar las capacidades .
- Crear una estructura en la alta dirección que impulse los ítems anteriores.

Todo esto se resume en su ciclo PDCA de mejora continua, conocido como *ciclo Deming* cuyo objetivo se basa en planificar (P), hacer lo que se planifica (D), medir y controlar lo que se hace (C), y actuar en consecuencia para mejorar los resultados (A), como se puede ver en la Figura 1.1.

[Juran \(2001\)](#) experto de la calidad y la gestión de la calidad. Declaró que la calidad es el conjunto de características que satisfacen las necesidades de los clientes, además calidad consiste el no tener deficiencias (Ver Figura 1.2). Propuso una trilogía de la calidad basada en:

- Planificación de la calidad
- Control de calidad
- Mejora de calidad

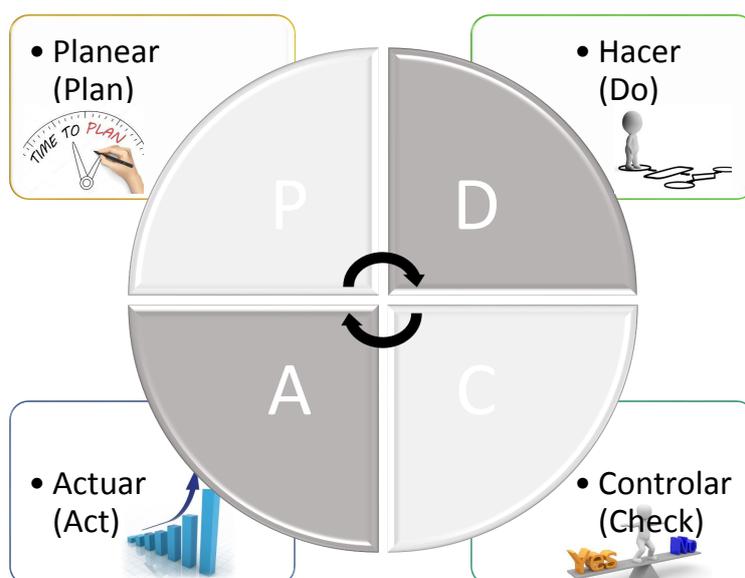


Figura 1.1: Ciclo PDCA - Ciclo Deming.

Fuente: Elaboración propia.

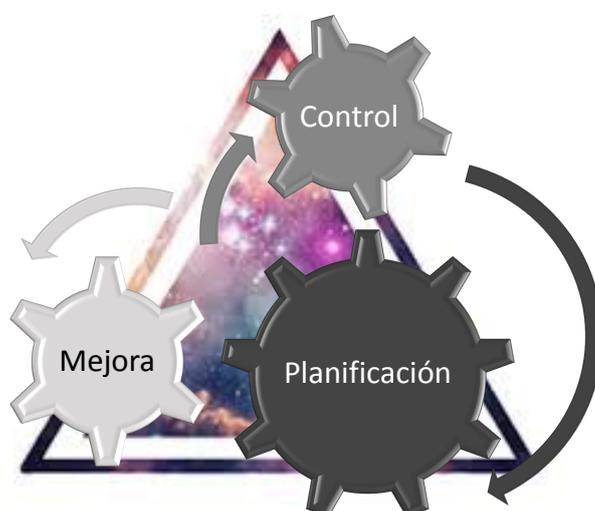


Figura 1.2: Trilogía de Juran.

Fuente: Elaboración propia.

Crosby (1996), empresario estadounidense, autor que contribuyó a la teoría gerencial y a las prácticas de la gestión de la calidad, la define como el cumplimiento de normas y

requerimientos precisos, añadiendo que la calidad es ajustarse a las especificaciones. Los 14 pasos para el mejoramiento de calidad de Crosby se muestran en la Figura 1.3, cuyo principal objetivo se basa en *cero* defectos y estos ayudan a lograrlo.

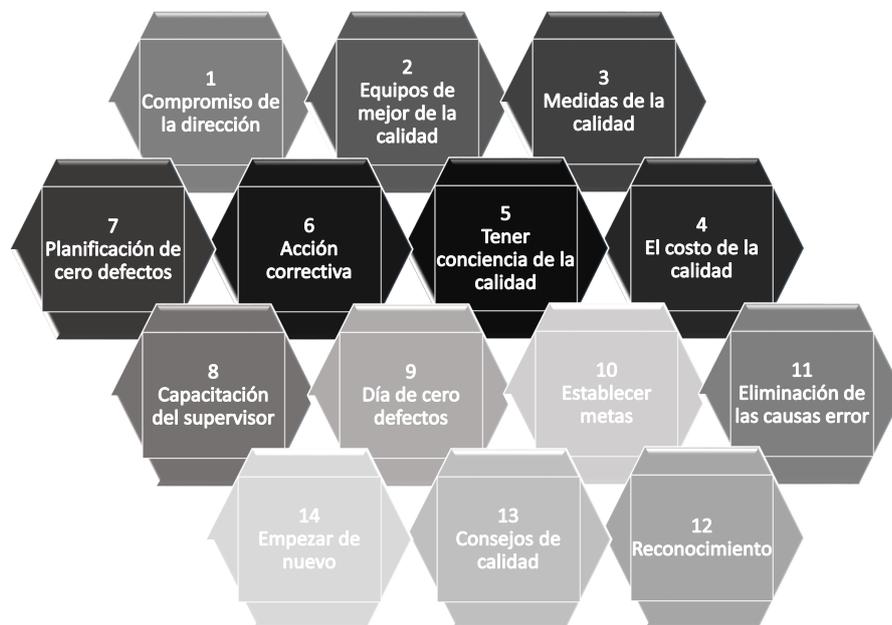


Figura 1.3: Catorce pasos de mejoramiento de calidad de Crosby.

Fuente: Elaboración propia.

En la actualidad, la real academia española define a la calidad como propiedad o conjunto de propiedades inherentes a algo, que permiten juzgar su valor.

## 1.2. Control estadístico de procesos

“No existe producto o servicio sin un proceso. De la misma manera, no existe proceso sin un producto o servicio”

[Harrington et al. \(1993\)](#)

La información suministrada por los gráficos de control de un proceso permiten identificar su estado, esto es, si está bajo control o fuera de control. En este último caso será necesario identificar las causas asignables que lo están afectando para tomar las acciones correctivas

pertinentes y lograr el estado de control. Para facilitar la comprensión del proceso, su posterior análisis y la detección de las causas asignables que lo afectan, la teoría del control estadístico de procesos proporciona herramientas de amplio uso y probada efectividad.

Si un producto va a cumplir o exceder las expectativas del cliente, generalmente debe ser producido por un proceso estable o repetible. Más precisamente, el proceso debe ser capaz de operar con poca variabilidad alrededor del objetivo o las dimensiones nominales de las características de calidad del producto. El control estadístico de procesos (SPC) es una poderosa colección de herramientas de resolución de problemas útiles para lograr la estabilidad del proceso y mejorar la capacidad mediante la reducción de variabilidad.

El control estadístico de procesos, es uno de los mayores desarrollos tecnológicos del siglo XX porque se basa en principios subyacentes sólidos, es fácil de usar, tiene un impacto significativo y se puede aplicar a cualquier proceso. Sus siete herramientas principales son:

1. Histograma
2. Hoja de verificación
3. Diagrama de Pareto
4. Diagrama de causa y efecto
5. Diagrama de concentración de defectos
6. Diagrama de dispersión
7. Diagrama de control

### **1.2.1. Histograma**

El histograma es un gráfico de barras verticales que representa la distribución de un conjunto de datos o bien la presentación de datos en forma ordenada con el fin de determinar la frecuencia con que algo ocurre. Éste muestra gráficamente la capacidad de un proceso, y si así se desea, la relación que guarda tal proceso con las especificaciones y las normas. También da una idea de la magnitud de la población y muestra las discontinuidades que se producen en los datos. Por su naturaleza gráfica, el histograma puede ayudar a identificar

e interpretar pautas que son difíciles de ver con una simple tabla de números y que son de poco valor si no aparecen suficientemente ordenados y clasificados. Además, se usan como herramienta de trabajo tanto para procesos industriales como dentro de las actividades habituales de gestión.

Los histogramas de frecuencias se usan con el objetivo de:

1. Conocer el grado con que se puede satisfacer los requisitos de calidad.
2. Establecer o modificar las especificaciones del artículo y el nivel de calidad aceptable (NCA).
3. Determinar los aspectos del proceso (ajuste y variabilidad) sobre los que se debe actuar para lograr la calidad deseada.
4. Conocer la presencia de causas asignables (ajenas del proceso), lo que permite detectar y tomar las medidas correctivas necesarias para eliminarlas.

Dos son, principalmente, las técnicas utilizadas en el análisis de la capacidad de un proceso: Histogramas y Gráficos de Control.

### **Análisis de la capacidad del proceso usando histogramas**

La distribución de frecuencias suele ser útil para estimar la capacidad del proceso<sup>1</sup>. Se requieren, por lo menos, entre 50/100 observaciones para realizar la estimación. Siempre que sea posible y antes de comenzar la recogida de datos, deben seguirse los siguientes pasos:

- a) Escoger adecuadamente la máquina o máquinas sobre las que va a realizarse el estudio de forma que sean representativas del conjunto de máquinas al que van a extenderse las conclusiones.
- b) Seleccionar las condiciones operativas del proceso definiendo cuidadosamente estas condiciones ya que su variación puede afectar a la capacidad del proceso.
- c) Seleccionar, cuando proceda, un operador representativo.

---

<sup>1</sup> Hervás (2015)

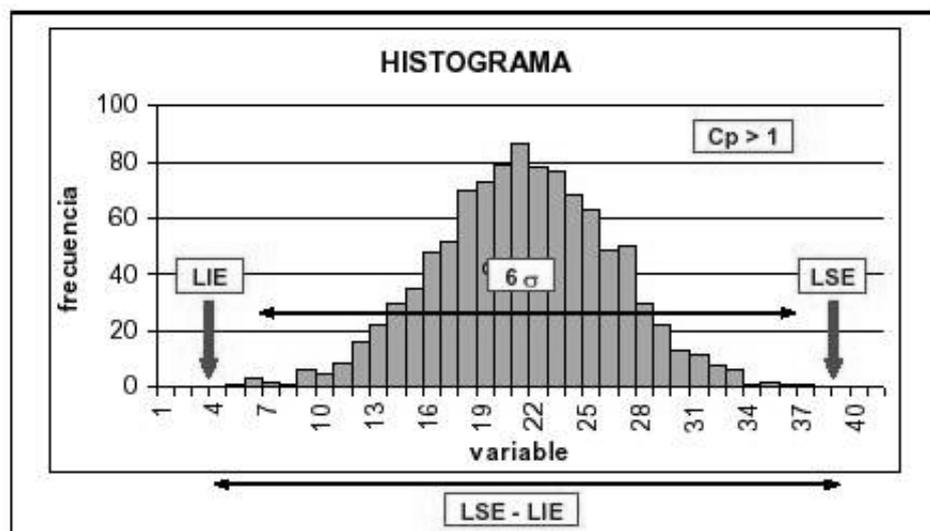
- d) Seguir estrechamente el proceso de toma de datos y registrar el orden en que se producen.

El histograma, con su valor medio  $\bar{X}$  y su desviación típica  $S$  estima la capacidad del proceso que es  $\bar{X} \pm 3S$ . Una ventaja del histograma es que nos da una impresión inmediata de cómo es la característica en estudio.

Para cuantificar la Capacidad de Proceso se utilizan coeficientes que permiten comparar el rango de especificaciones con la fluctuación natural del proceso. Uno de ellos es  $C_p$ :

$$C_p = \frac{LSE - LIE}{6\sigma},$$

donde LSE y LIE son, respectivamente, el Límite Superior y el Límite Inferior de Especificación. Si el proceso tiene capacidad para fabricar el producto, entonces  $C_p > 1$ . En general se exige  $C_p > 1.30$  para mayor seguridad.



#### 1.2.1.1. Ventajas

Entre las ventajas que presenta el histograma tenemos:

1. Proporciona, mediante el estudio de la distribución de los datos, un excelente punto de partida para generar hipótesis acerca de un funcionamiento insatisfactorio.

2. Es un medio eficaz para transmitir a otras personas información sobre un proceso de forma precisa e inteligible.
3. Es especialmente útil cuando se tiene un amplio número de datos que es preciso organizar para analizar más detalladamente o tomar decisiones sobre la base de ellos.
4. Su construcción ayudará a comprender la tendencia central, dispersión y frecuencias relativas de los distintos valores.

#### **1.2.1.2. Desventajas**

Entre las desventajas que presenta el histograma tenemos:

1. No es el gráfico más apropiado para representar el cambio de una variable estadística a través del tiempo (son los gráficos o diagramas de líneas).
2. No identifica la causa real o principal de un problema o efecto.

#### **1.2.2. Hoja de Verificación**

La hoja de verificación es también llamada hoja de control o de chequeo y consiste en un impreso con formato de tabla o diagrama, destinada a registrar y compilar datos mediante un método sencillo y sistemático, como la anotación de marcas asociadas a la ocurrencia de determinados sucesos. Esta técnica de recolección de datos se prepara de manera que su uso sea fácil e interfiera lo menos posible con la actividad de quien realiza el registro, (ver Figura 1.4).

Las hojas de control se usan para:

- a. Verificar o examinar artículos defectuosos.
- b. Examinar o analizar la localización de defectos.
- c. Verificar las causas de defectivos.
- d. Verificación y análisis de operaciones

*Hoja de verificación de control de calidad*

 Nombre     
  Dirección     
  Fecha de recogida de datos

Tipo de defecto/ ocurrencia del evento	Datos														Total
	Lunes		Martes		Miércoles		Jueves		Viernes		Sábado		Domingo		
	Error	Acción	Error	Acción	Error	Acción	Error	Acción	Error	Acción	Error	Acción	Error	Acción	
Descripción 1															
Descripción 2															
Descripción 3															
Descripción 4															
Descripción 5															
Descripción 6															
Total															

 Observaciones:
   
\_\_\_\_\_
   
\_\_\_\_\_

Figura 1.4: Hoja de verificación.

### 1.2.2.1. Ventajas

Entre las ventajas que presenta la hoja de control tenemos:

1. Supone un método con el que se obtienen datos fáciles de comprender y que son obtenidos mediante un simple y eficiente formato de tabla o diagrama que puede ser aplicado a cualquier área de la organización.
2. Reflejan rápidamente las tendencias y patrones subyacentes en los datos.
3. Es utilizada tanto en el estudio de los síntomas de un problema, como en la investigación de las causas o en la recopilación y análisis de datos para probar alguna hipótesis.
4. Se usa como punto de partida para la elaboración de otras herramientas, como por ejemplo los gráficos de control.

### 1.2.2.2. Desventajas

Entre las desventajas que presenta la hoja de control tenemos:

1. De acuerdo al enfoque que se le dé al formato puede pasar de ser fácil y eficiente a uno no funcional.
2. Su uso excesivo puede llevar a la obtención de datos sin ningún objetivo concreto e importante.

### 1.2.3. Diagrama de Pareto

El diagrama de Pareto, también llamado curva cerrada o Distribución A-B-C, es un método gráfico de análisis en donde se organizan diversas clasificaciones de datos por orden descendente, de izquierda a derecha por medio de barras sencillas, que permite separar entre los efectos más importantes de un problema de los que son la minoría, de modo que se pueda asignar un orden de prioridades. [Juran et al. \(1983\)](#) aplicó este concepto a la calidad, obteniéndose lo que hoy se conoce como la regla 80/20, la cual nos dice que si se tiene un problema con muchas causas, podemos decir que el 20 % de las causas resuelven el 80 % del problema y el 80 % de las causas sólo resuelven el 20 % del problema, el esquema de este concepto es presentado en la Figura 1.5.

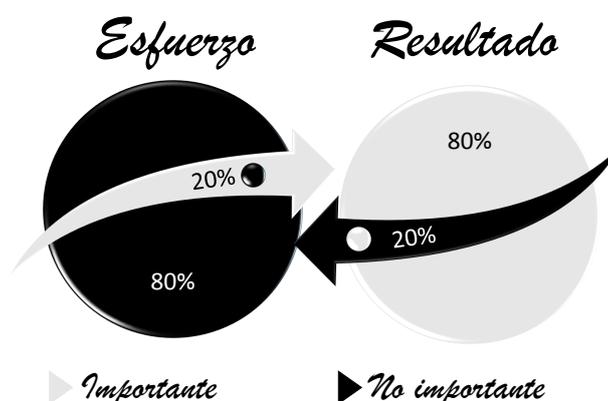


Figura 1.5: Ley de Pareto.

La gráfica es útil al permitir identificar visualmente en una sola revisión tales minorías de características vitales a las que es importante prestar atención y de esta manera utilizar todos los recursos necesarios para llevar a cabo una acción correctiva sin malgastar esfuerzos, Sales (2002). Los diagramas de Pareto están basados en el principio de que la mayoría de los efectos de un proceso son el resultado de sólo unas cuantas causas. El diagrama identifica las pocas causas críticas que están generando la mayoría de efectos en el proceso, a fin de dirigir la atención y el análisis a éstos. Además, lo constituyen histogramas de frecuencias absolutas y relativas de la ocurrencia del efecto en función de los factores y causas que lo generan, así como el gráfico de frecuencia absoluta acumulada y frecuencia relativa acumulada.

#### **1.2.3.1. Ventajas**

Entre las ventajas que presenta el diagrama de Pareto tenemos:

1. Identificar oportunidades para la mejora y pueden ser asimismo utilizados tanto para investigar efectos como problemas.
2. Identifica los principales efectos y establece las prioridades de las soluciones.
3. Es un método que brinda resultados fáciles de comprender y los cuales se obtienen mediante un proceso simple y eficiente que se puede aplicar en cualquier área de una organización.
4. Las decisiones que se toman se basan en datos y hechos objetivos, no en datos o ideas subjetivas.

#### **1.2.3.2. Desventajas**

Entre las desventajas que presenta el diagrama de Pareto tenemos:

1. Determina cuál es el efecto principal de un problema y no la causa principal que lo provoca.
2. La selección, confiabilidad y consistencia de los datos va a depender de la de la persona que haga el estudio.

3. No da una idea clara sobre las causas de raíz y es necesario más del 50 % de las categorías para sumar un 60 % del efecto de calidad

#### 1.2.4. Diagrama Causa - Efecto

Los diagramas causa–efecto, conocidos también como diagrama de Ishikawa (por su creador, [Ishikawa \(1943\)](#)) o diagramas espina de pescado, por la forma en que se van colocando cada una de las causas o razones que originan un problema, permiten identificar las causas que están generando efectos perturbadores del proceso. Este diagrama es una forma de organizar y representar las diferentes teorías propuestas sobre las causas de un problema y bien ayuda a graficar las causas del problema que se estudia para analizarlas. El diagrama debe de incluir los siguientes elementos:

- a. El problema principal que se desea analizar, el cual se coloca en el extremo derecho del diagrama. Se aconseja encerrarlo en un rectángulo para visualizarlo con facilidad.
- b. Las causas principales que a nuestro entender han originado el problema.

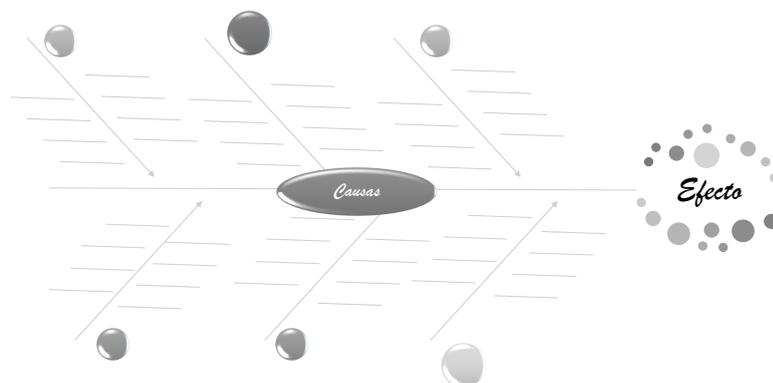


Figura 1.6: Diagrama de Causa - Efecto.

Fuente: Elaboración propia.

La Figura 1.6 está constituida por un eje central horizontal que es conocida como “línea principal o espina central”, posee varias flechas inclinadas que se extienden hasta el eje

central, al cual llegan desde su parte inferior y superior, según el lugar donde se haya colocado el problema que se estuviera analizando o descomponiendo en sus propias causas o razones, cada una de ellas representa un grupo de causas que inciden en la existencia del problema, cada una de estas flechas a su vez son tocadas por flechas de menor tamaño que representan las “causas secundarias” de cada “causa” o “grupo de causas del problema”.

#### **1.2.4.1. Ventajas**

Entre las ventajas que presenta el diagrama causa–efecto, tenemos que:

1. Sirve para que la gente conozca con profundidad el proceso con que trabaja, visualizando con claridad las relaciones entre los efectos y sus causas.
2. Sirve para guiar las discusiones, al exponer con claridad los orígenes de un problema de calidad y permite encontrar más rápidamente las causas asignables cuando el proceso se aparta de su funcionamiento habitual.
3. Permite visualizar de una manera muy rápida y clara, la relación que tiene cada una de las causas con las demás razones que inciden en el origen del problema.

#### **1.2.4.2. Desventajas**

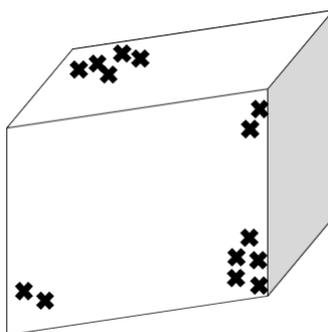
Entre las desventajas que presenta el diagrama causa–efecto, tenemos:

1. Variación del grupo al momento que este se concentre analizando el contenido del problema llegando a enfocar la historia del problema y los distintos intereses personales de los integrantes del equipo.
2. Fallo al momento de clasificar y relacionar las interacciones entre factores que están afectando al resultado de un proceso.
3. Se puede dar una mala asignación a las causas principales del problema, lo cual generaría un mal análisis, el cual implica la pérdida de recursos.

### 1.2.5. Diagrama de concentración de defectos

Un diagrama de concentración de defectos es una imagen de la unidad, que muestra todas las vistas relevantes. Luego, se dibujan los diversos tipos de defectos en la imagen, y se analiza el diagrama para determinar si la ubicación de los defectos en la unidad transmite información útil sobre las posibles causas de los defectos. Los diagramas de concentración de defectos son muy útiles en el paso de análisis de DMAIC (es un procedimiento estructurado de resolución de problemas ampliamente utilizado en la calidad y la mejora de procesos. Por sus siglas en inglés, DMAIC significa: Definir, Medir, Analizar, Mejorar y Controlar). Cuando los datos de defectos se representan en un diagrama de concentración de defectos en un número suficiente de unidades, los patrones surgen con frecuencia y la ubicación de estos patrones a menudo contiene mucha información sobre las causas de los defectos. Los diagramas de concentración de defectos son herramientas importantes para la resolución de problemas en muchas industrias, incluidas las operaciones de enchapado, pintura y revestimiento, operaciones de fundición y fundición, mecanizado y ensamblaje electrónico. El diagrama de concentración (Ruiz et al. (2016)) nos da una respuesta tridimensional e irrefutable a la pregunta:

#### 1.2.5.1. ¿Dónde ocurren los defectos?



Preguntas: “ ¿En qué área ocurre el defecto?”y te dicen “En todas ellas”pero al recibir una imagen de las áreas pides que marquen con una cruz en la respectiva área cada vez que salga mala y después de determinado tiempo se obtienen los siguiente resultados, lo cual se evidencia por medio de la Figura 1.7.

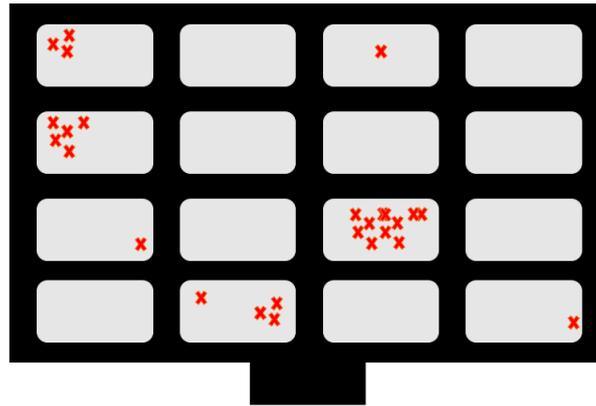


Figura 1.7: Áreas con defectos en determinado tiempo.

### 1.2.6. Diagrama de dispersión

El diagrama de dispersión, también conocido como gráfico de correlación, permite básicamente comprender si se encuentran vinculadas entre sí dos magnitudes y en qué medida; además, estudia la intensidad de la relación de éstas (Périsse (2007)). Dadas dos variables  $X$  y  $Y$ , se dice que existe una correlación entre ambas si éstas son directa o inversamente proporcionales (correlación positiva o negativa). En un gráfico de dispersión (Figura 1.8) se representa cada par  $(x, y)$  como un punto donde se cortan las coordenadas de  $X$  e  $Y$ , por lo que se convierte en una representación gráfica en un eje de coordenadas de los datos que se recogen sobre dos variables para poder estudiar si existe relación de causa efecto entre ellas. Sirve para verificar causas reales, definir y medir relaciones existentes entre dos variables.

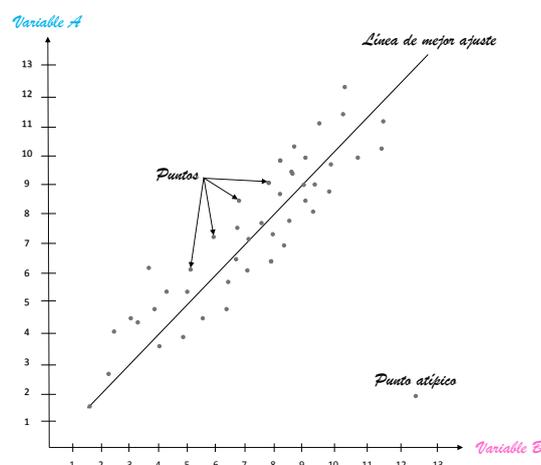


Figura 1.8: Diagrama de dispersión.

Se pueden interpretar varios tipos de correlación, como se logra visualizar a través de los patrones mostrados en los diagramas de dispersión en la Figura 1.9 (Ojeda (2015)). Estos son<sup>2</sup>:

1. Positivo: los valores aumentan juntos.
2. Negativo: un valor disminuye a medida que el otro aumenta
3. Nulo: no existe correlación
4. Lineal: existe relación directa en los valores
5. Exponencial: tiene un ritmo que aumenta cada vez más rápidamente
6. En forma de U, que no hay un patrón

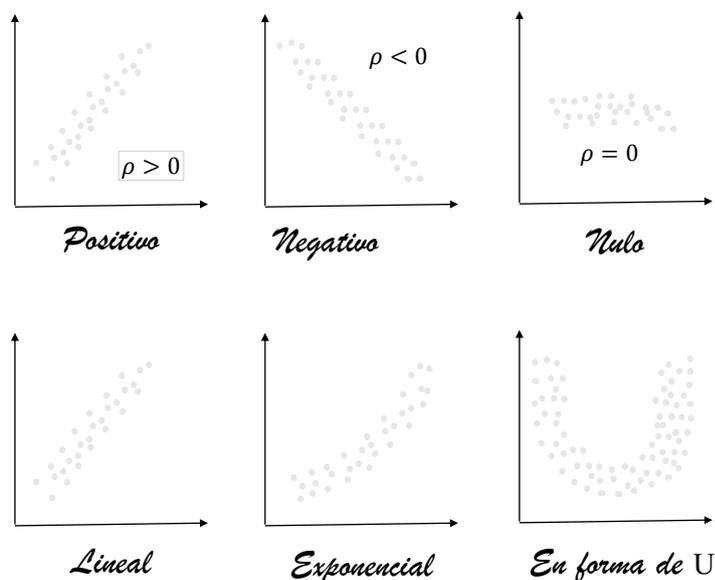


Figura 1.9: Tipos de correlación.

La fuerza de la correlación puede determinarse por la cercanía en los cambios de ambas variables. A mayor correlación, más tienden a formar una línea recta en el diagrama de dispersión. Los puntos que terminan muy lejos del conjunto general de puntos se conocen como valores atípicos. Las líneas o curvas se ajustan dentro del gráfico para ayudar en

<sup>2</sup>Neria (2014)

el análisis y se dibujan tan cerca de todos los puntos como sea posible para mostrar cómo se condensaron todos los puntos en una sola línea. Esto se conoce normalmente como “línea de mejor ajuste” o “línea de tendencias” y se puede utilizar para hacer estimaciones mediante interpolación. Los diagramas de dispersión son ideales cuando se tienen datos numéricos emparejados y se desea ver si una variable afecta a la otra. Sin embargo, recuerde que la correlación no es causal y otra variable inadvertida puede estar influyendo en los resultados.

### 1.2.7. Diagrama de control

Los diagramas o cartas de control, son una de las principales herramientas de calidad, porque permiten monitorear los procesos de forma tal que posibilitan su estandarización, la estabilidad, la evaluación de la mejora o del retroceso del mismo y también la predicción del futuro inmediato según las tendencias actuales de las variables objeto de estudio. Las cartas de control funcionan con dos clases de variación, las cuales se denominan comunes y especiales (o asignables). El propósito fundamental de los diagramas de control, es justamente ayudarnos a distinguir entre las dos situaciones por medio de diagramar series de tiempo en las que el investigador en el proceso hace observaciones de la variable de interés a través del tiempo (comportamiento dinámico)<sup>1</sup>.

## 1.3. Carta de control

### 1.3.1. Definición

De acuerdo con [Shewhart \(1924\)](#), una carta de control es una herramienta estadística usada principalmente para el estudio, control y mejora de procesos repetitivos. Existen cartas de control univariadas y multivariadas, las primeras se enfatizan en el estudio de una característica de calidad y es conveniente monitorear tanto la media como la varianza del proceso, mientras que en las cartas de control multivariadas, hay más de una característica de calidad en el proceso y de manera análoga, es conveniente el monitoreo del vector de medias y de la matriz de covarianzas. El objetivo de las cartas de control es observar

---

<sup>1</sup>[Coronado \(2007\)](#)

y analizar mediante el uso de datos estadísticos la variabilidad del proceso de interés a través del tiempo ([Gutiérrez y De la Vara \(2004\)](#)). Por otro lado, una carta de control es un diagrama que sirve para examinar si un proceso se encuentra en una condición estable, o para asegurar que se mantenga en esa condición. En caso contrario, si el proceso se encuentra fuera de control el objetivo es traerlo a control. También puede verse como la prueba continua de las hipótesis, en efecto:

$$\begin{cases} H_0 : \text{El proceso se encuentra bajo control estadístico} \\ H_1 : \text{El proceso se encuentra fuera control estadístico .} \end{cases}$$

Para las cartas de control se pueden construir intervalos de confianza, las cuales nos proporcionan información sobre la estabilidad del proceso. El proceso será estadísticamente estable si el 95 % de las veces, el resultado está dentro del intervalo construido, esto, considerando la proporción de falsa alarma de 5 %.

La estructura general de las cartas de control contiene una *línea central* (LC), una línea superior que marca el *límite de control superior* (LCS), y una línea inferior que marca el *límite de control inferior* (LCI). Los puntos contienen información sobre las lecturas hechas; pueden ser promedios de grupos de lecturas, o sus rangos, o bien las lecturas individuales mismas. Los límites de control marcan el intervalo de confianza en el cual se espera que caiga la mayor parte de los puntos si el proceso está bajo control (ver Figura 1.10). Algunas cartas de control pueden tener un solo límite de control (por lo general, el límite superior).

Una carta de control, entre otras cosas sirve como ayuda para:<sup>1</sup>

- Determinar el estado de control de un proceso.
- Observar el comportamiento de un proceso en el tiempo.
- Indicar si un proceso ha mejorado o ha empeorado.
- Identificar las dos fuentes de variación de un proceso.
- Detectar problemas.

---

<sup>1</sup>SPC (2013)

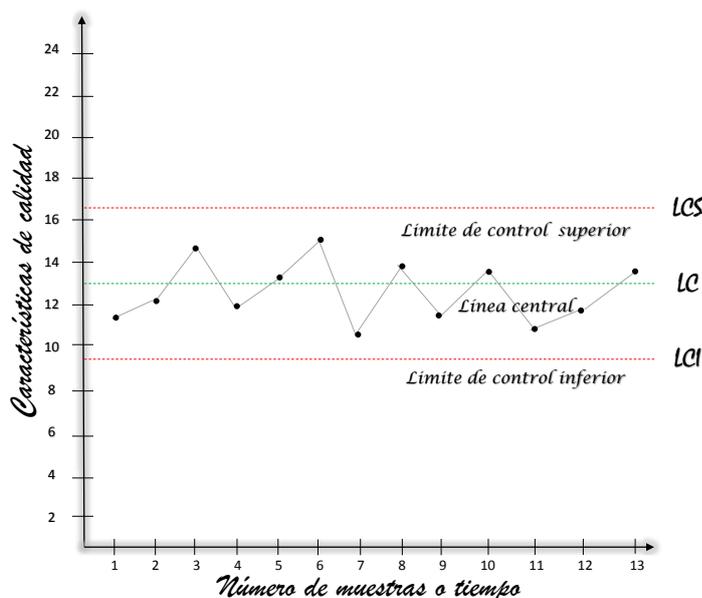


Figura 1.10: Estructura de un gráfico de control.

Fuente: Elaboración propia.

### 1.3.2. Tipos de variabilidad

Mediante el uso de las cartas de control se pretende identificar las principales fuentes de variación del proceso, las cuales se identifican como: variabilidad no asignable y variabilidad asignable.<sup>2</sup>

#### 1.3.2.1. Variabilidad debida a causas no asignables (comunes o aleatorias)

Variabilidad que aparece de manera natural en el proceso debida al azar y es inherente a la calidad, se da por la forma en que se diseña y opera normalmente el sistema, y estas causas son aquellas que son parte inherente del proceso o sistema y que afectan a todos los que están trabajando dentro de él.

La variación debida a estas causas es aleatoria y refleja las variaciones de todo tipo que proviene del diseño del proceso o sistema y de su forma normal de operar. La clave es que no es posible averiguar la causa de una variación debida a causas comunes. Esta variación aleatoria proviene de un número muy grande de interacciones de todos los factores,

<sup>2</sup>Rivera (2011)

tangibles e intangibles, que intervienen normalmente en el proceso. Entonces, cuando el proceso solamente se ve afectado por causas comunes, se dice que es estable, o que está en estado de control estadístico.

### 1.3.2.2. Variabilidad debida a causas asignables (especiales o no aleatorias)

Variabilidad originada por circunstancias o situaciones especiales ajenas al proceso. Este tipo de variabilidad a menudo puede ser identificada y eliminada del proceso. Las causas especiales, por su naturaleza relativamente discreta, a menudo pueden ser identificadas y eliminadas si se cuenta con los conocimientos y condiciones para ello, [Bergengruen \(2013\)](#). Es algo especial, no es parte del sistema de causas comunes, esta variación es causada por situaciones o circunstancias especiales que no están presentes permanentemente en el sistema; por ejemplo, una falla ocasionada por el mal funcionamiento de una pieza en una máquina o el empleo de materiales no habituales. Un proceso en el que están presentes causas especiales de variación se dice que está fuera de control estadístico (o es inestable). Este tipo de procesos es impredecible en el futuro inmediato porque en cualquier momento pueden aparecer esas situaciones que tienen un efecto especial sobre la variabilidad. La Figura 1.11 muestra la probabilidad y causas de variación asignables:

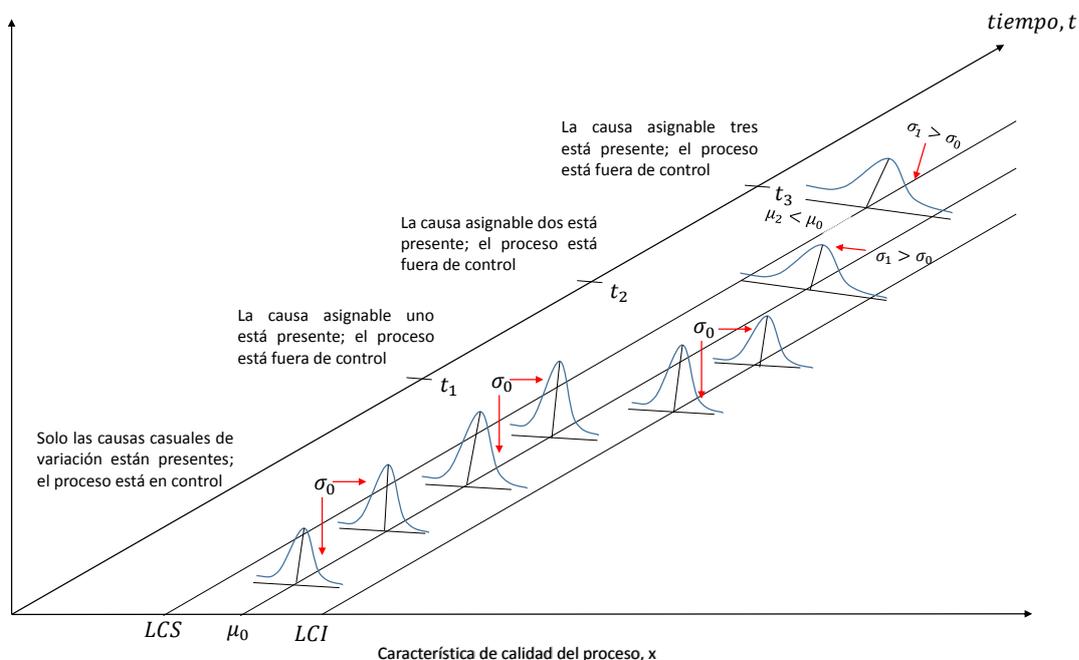


Figura 1.11: Probabilidad y causas de variación asignables.

### 1.3.3. Construcción de una carta de control

Un gráfico o carta de control tiene una línea central que representa el valor medio de la característica de calidad cuando el proceso está bajo control. Además se consideran otras dos líneas horizontales llamadas límite de control superior (LCS) y límite de control inferior (LCI). En términos generales, un gráfico de control es una representación gráfica en los ejes cartesianos, donde en el eje horizontal se indica el número de la muestra o el tiempo en que se obtiene, y en el eje vertical se indican los valores observados en las muestras. El modelo general de las cartas tipo Shewhart es:

$$LCI = \mu_T - k\sigma_T,$$

$$LC = \mu_T,$$

$$LCS = \mu_T + k\sigma_T,$$

donde:

- $T$  es un estadístico muestral que mide cierta característica en la que se tiene interés,
- $\mu_T$  es la media de  $T$ ,
- $\sigma_T$  es la desviación estándar  $T$ ,
- $k$  es una constante conocida, que indica la distancia de los límites de control a la línea central expresada en unidades de desviación típica.

Los límites de control se escogen de tal manera que si el proceso está bajo control, casi la totalidad de los puntos muestrales caigan entre ellos. Para establecer los límites, se puede:

1. fijar el valor de  $k$  y luego determinar la probabilidad de cometer error tipo I,
2. ó fijar la probabilidad de cometer error tipo I, y con base en ello, determinar el valor de  $k$

Según [Shewhart \(1924\)](#), la forma genérica de una carta de control, desde un punto de vista analítico y desde un punto de vista gráfico, es el que se puede ver en la Figura 1.12

La posición de los límites de control se determina a partir del valor de la constante  $k$  y se fija un  $\alpha$  de acuerdo al error tipo I. Normalmente se toma  $k = 3$ , en cuyo caso los

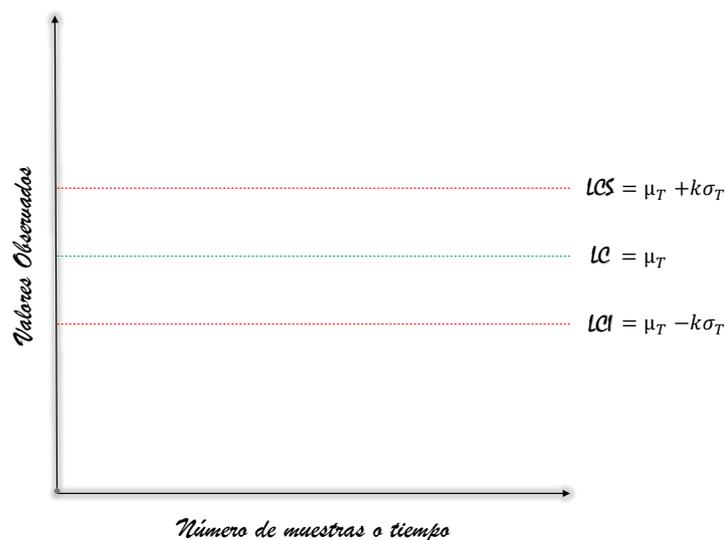


Figura 1.12: Carta de control genérica.

límites se denominan límites de control  $3\sigma$ , o límites de acción. Cuando la característica de calidad que se está estudiando sigue una distribución normal, la probabilidad de error de tipo I es de  $\rho = 0.0027$ .

Una gráfica de control<sup>1</sup> sigue la evolución de características de un proceso (continuas o discretas), que se desean mantener bajo control, lo que garantizará que el producto final cumpla con los estándares de calidad. Por lo tanto, es necesario determinar la distribución de probabilidad que tengan las variables, ya que ello ayudará a establecer los límites de control y obtener la medida de rendimiento del gráfico, [Montgomery y Wadsworth \(1972\)](#). Por lo general, si la variable a controlar es continua, lo ideal sería que esta tenga una distribución de probabilidad normal  $N(\mu, \sigma^2)$ , ya que muchos estadísticos se basan en ese supuesto. Si ello no se cumple, se puede utilizar alguna transformación matemática o el Teorema del Límite Central, para aproximar la normalidad, [García \(2014\)](#). En la Tabla 1.1 se presentan los parámetros más comunes a controlar y sus estimadores muestrales. Los gráficos que controlan los primeros cuatro parámetros, utilizan estadísticos muestrales, cuya muestra se supone deben pertenecer a una población normalmente distribuida. En el caso del estadístico muestral  $\bar{X}$ , el teorema de límite central resolvería el problema<sup>2</sup>.

<sup>1</sup>[Plaza \(2018\)](#)

<sup>2</sup>[García \(2014\)](#)

Parámetro	Estimador
Media ( $\mu$ )	$\bar{X} = \frac{1}{n} \sum_{i=1}^n x_i$
Rango	$R = \text{Max}(x_i) - \text{Min}(x_i)$
Varianza ( $\sigma^2$ )	$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$
Desviación estándar ( $\sigma$ )	$S = \sqrt{S^2}$
Proporción de unidades no conformes por lote ( $p$ )	$\hat{p} = \frac{D}{n}$
Proporción de defectos por lote	$u = \frac{x}{n}$ siendo $x$ el número de defectos totales en la muestra
Unidades no conformes en el lote	Unidades no conformes en la muestra ( $D$ )
Cantidad de defectos por lote	Cantidad de defectos en la muestra ( $c$ )

Tabla 1.1: Parámetros más comunes a controlar y sus estimadores.

Fuente: Control multivariante estadístico de variables discretas tipo Poisson, 2014.

### 1.3.4. Criterios para concluir cuando un proceso está fuera de control

#### 1.3.4.1. ¿Cuándo un proceso está bajo control?

Un proceso está bajo control estadístico cuando no hay causas asignables presentes. Esto es equivalente a decir que el proceso permanezca estable, es decir que los parámetros de

la distribución permanezcan invariables y por lo tanto puede realizarse una predicción del intervalo en el que se encontrarán los valores de la característica de respuesta. Por lo tanto, debe tratar de conocerse todo lo que sea posible de los fundamentos tecnológicos del proceso, ya que puede dar pistas sobre el tipo de distribución que seguirán los datos. En ningún caso debe “darse la normalidad por supuesta”. Debe comprobarse y en caso de que los datos no sean normales, deben aplicarse métodos especiales (Rojas (2006)). Por otro lado, Carro y González Gómez (2012) dicen que un proceso está bajo control estadístico cuando la localización, expansión o forma de sus distribución no cambia con el tiempo.

#### 1.3.4.2. ¿Cuándo un proceso está fuera de control?

Un proceso se considera fuera de control, cuando hay causas asignables presentes. En la práctica se asume que esto puede suceder cuando al menos uno de los valores que se gráfica en el diagrama de control cae fuera de los límites de control. Sin embargo, este no es el único criterio, existen la reglas de Western Electric.

En la práctica, cuando esto sucede se debe investigar a qué se debe esta anomalía o si solo se trata de una falsa alarma. En consecuencia, los diagramas de control indican que el proceso está fuera de control cuando los puntos muestrales se encuentran fuera de los límites de control, o bien cuando presentan una distribución no aleatoria en torno a la línea central del diagrama. Por ejemplo, la presencia de puntos muestrales consecutivos que describen una poligonal creciente o decreciente se considera un indicio de no aleatoriedad. Estas rachas de puntos muestrales también se pueden situar uniformemente por encima o por debajo de la línea central. Otro ejemplo de no aleatoriedad se presenta cuando todos los puntos muestrales del diagrama se sitúan por encima o por debajo de la línea central. El manual Western Electric sugirió un conjunto de reglas de decisión para detectar patrones no aleatorios en las cartas de control. Específicamente, sugiere concluir que el proceso está fuera de control si:

1. Uno o más puntos fuera de los límites de control de tres sigma.
2. Dos de tres valores consecutivos caen fuera de los límites de advertencia de dos sigma del mismo lado.

3. Cuatro de cinco puntos consecutivos se trazan a una distancia de un sigma o más allá de la línea central
4. Ocho puntos consecutivos trazan en un lado de la línea central.

Algunas reglas de sensibilización para los gráficos de control de Shewhart, adicionales a las antes mencionadas son:

5. Seis puntos seguidos aumentando o disminuyendo constantemente.
6. Quince puntos seguidos en la zona C (tanto arriba como debajo de la línea central).
7. Catorce puntos seguidos alternando arriba y abajo.
8. Ocho puntos seguidos a ambos lados de la línea central sin ninguno en la zona C.
9. Un patrón inusual o no aleatorio en los datos.
10. Uno o más puntos cerca de un límite de advertencia o control

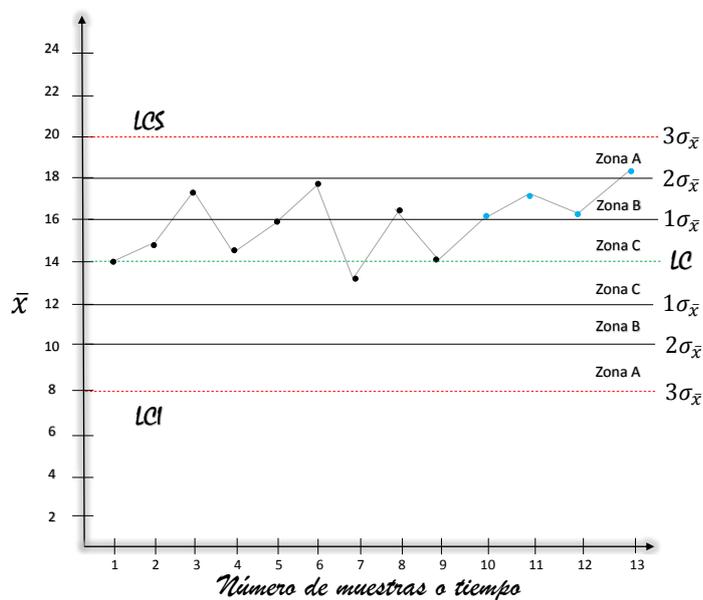


Figura 1.13: Reglas de Western Electric.

La Figura 1.13 muestra una carta de control con los límites de uno, dos y tres sigmas utilizados en el procedimiento de Western Electric (donde los últimos cuatro puntos muestran una violación de la regla 3). Estos límites dividen la tabla de control en tres zonas (A, B y C) a cada lado de la línea central. En consecuencia, las reglas de Western Electric a veces se denominan reglas de zona para gráficos de control. Los últimos cuatro puntos caen en la zona B o más allá, por lo tanto, dado que cuatro de los cinco puntos consecutivos exceden el límite de un sigma, el procedimiento de Western Electric concluirá que el patrón no es aleatorio y el proceso está fuera de control.

### 1.3.5. Probabilidad de falsa alarma ( $\alpha$ )

Una falsa alarma es un punto que cae fuera de los límites de control, cuando realmente el proceso no ha cambiado, es decir, no ha salido de control. Esto quiere decir que si observamos un punto fuera de los límites de control de alguno de los gráficos de control, podemos tener dos escenarios:

- a) El proceso no se ha salido de control y el punto representa una falsa alarma
- b) El proceso realmente se ha salido de control.

La mayor parte de la literatura contemplan la posibilidad *b*), pero no se puede perder de vista que la posibilidad *a*) teniendo en cuenta que las cartas se diseñan de tal manera que si el proceso está bajo control es “poco probable” que un punto se salga de los límites de control, en caso contrario se supone es que el proceso no está bajo control.

### 1.3.6. Longitud de corrida (RL)

La longitud de corrida (RL), es una variable aleatoria y corresponde al número de puntos graficados en la carta hasta que un punto se sale de los límites. Al promedio de todas las longitudes de corrida ( $E(RL)$ ) se le denomina ARL y se espera que si el proceso está bajo control, la ARL sea grande.

### 1.3.6.1. ARL

Montgomery (2007), define la Longitud Promedio de Corrida (ARL) como el número promedio de puntos que deben ser graficados antes de que un punto indique una condición fuera de control. Si las observaciones del proceso no están correlacionadas, entonces, para cualquier gráfico de control de Shewhart, puede ser calculada de la siguiente manera:

$$ARL = \frac{1}{\rho}, \quad (1.1)$$

donde  $\rho$  es la probabilidad de que alguno de los puntos graficados supere los límites de la carta de control. En general cuando no se posee mucha información del proceso, es recomendable usar un  $\rho = 0.0027$  debido a que con este valor se tiene límites aproximados de  $3\sigma$ .

### 1.3.6.2. $ARL_0$

El  $ARL_0$ , es un concepto relacionado con procesos bajo control, los cuales tienen causas fortuitas (variabilidad natural del proceso). Está relacionado con el **Error Tipo I** ( $\alpha$ ), donde  $\alpha$  es la probabilidad de obtener falsas alarmas.

$$ARL_0 = \frac{1}{\alpha}. \quad (1.2)$$

### 1.3.6.3. $ARL_1$

El  $ARL_1$ , es un concepto relacionado con procesos fuera de control, los cuales tienen causas asignables que afectan en gran medida el proceso. Está relacionado con el **Error Tipo II** ( $\beta$ ), donde  $\beta$  es la probabilidad de que suceda el Error Tipo II (decir que el proceso está bajo control, cuando en realidad está fuera de control).

$$ARL_1 = \frac{1}{\beta - 1}. \quad (1.3)$$

**Nota:** El valor del  $ARL_1$  significa el número promedio de puntos que se va a demorar para saber que el proceso está fuera de control.

### 1.3.7. Tipos de cartas de control

De acuerdo con el tipo de variable que representa a la característica de calidad de interés, existen dos tipos de cartas de control, las cuales son: de Variables y por Atributos, los cuales se pueden observar en la Tabla 1.2.

	Carta $\bar{X} - R$	Promedios y Rangos
De	Carta $\bar{X} - s$	Promedios y Desviación Estándar
Variables	Carta $X - R$	Medianas y Rangos
	Carta $x - R$	Lecturas Individuales y Rangos
Por	$p$	Porcentaje de unidades o procesos defectuosos
	$np$	Número de unidades o procesos defectuosos
	$c$	Número de defectos por área de oportunidad
	$u$	Porcentaje de defectos por área de oportunidad

Tabla 1.2: Clasificación de las cartas de control.

Fuente: Elaboración propia.

#### 1.3.7.1. Carta de control para variables

Estas cartas se aplican a características de calidad de tipo continuo, que intuitivamente son aquellas que requieren un instrumento de medición (peso, volúmenes, voltaje, longitudes, resistencia, temperatura, etc). Las más usuales son X (medias), R (Rango),  $T^2$ ,  $S^2$ .

#### 1.3.7.2. Carta de control para atributos

Existen características de calidad de un producto que no son evaluadas con un instrumento de medición de escala continua o en una escala numérica. En estos casos, el producto se juzga como conforme o no conforme y así se determina el número de defectos o items defectuosos. Este tipo de características de calidad son monitoreadas a través de las cartas de control por atributos: P (proporción o fracción de artículos defectuosos), NP (número de unidades defectuosas), C (número de defectos) o U (número de defectos por unidad)

La diferencia fundamental radica en la forma en que usan la información. Además de las cartas mencionadas, que pertenecen a las cartas tipo Shewhart, existen otros grupos de

cartas de control, con las cuales se pretende detectar más rápido un cambio en el proceso y reducir la frecuencia de falsas alarmas. Entre estas cartas las más conocidas son las cartas EWMA y CUSUM.

### 1.3.8. Estimación de los parámetros en una carta de control

Muchos estudios tales como [Blanco \(2012\)](#) y [Willem y Wilbert \(2004\)](#) han mostrado que el desempeño de las cartas de control depende, entre otras cosas, del conocimiento o no de los parámetros asociados a éstas. Cuando los parámetros no son conocidos, sino que hay que estimarlos a partir de una muestra preliminar de observaciones, se ha encontrado que en algunos casos se hace necesario utilizar muestras preliminares muy grandes para garantizar un desempeño similar al que la teoría establece cuando los parámetros son conocidos.

Para los parámetros más conocidos, por ejemplo la media ( $\mu$ ) y la varianza ( $\sigma^2$ ), usualmente se estiman con el estimador de mínimos cuadrados. En la práctica, deben estimarse a partir de la muestra o subgrupos preliminares tomados cuando se considera que el proceso está bajo control. En general estas estimaciones deben basarse en al menos 20 o 25 muestras.

Supongamos que contamos con  $m$  muestras, cada una de las cuales tiene el mismo tamaño muestral  $n$ . Sean  $\bar{x}_1, \bar{x}_2, \bar{x}_3, \dots, \bar{x}_m$  los promedios de cada muestra. Entonces el mejor estimador de  $\mu$ , el promedio de la población, es el gran promedio,  $\bar{\bar{x}}$ .

$$\hat{\mu} = \bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \dots + \bar{x}_m}{m},$$

y la estimación de la desviación estándar  $\sigma$  puede hacerse a partir de las desviaciones estándar o bien de los rangos de las muestras. Cuando los tamaños muestrales son menores que 8, el método del rango proporciona una buena estimación. Para tamaños muestrales mayores que 8, se reemplazará por el método de las desviaciones estándar.

Si  $x_1, x_2, \dots, x_n$  es una muestra de tamaño  $n$ , entonces el rango de la muestra es la diferencia entre las observaciones menor y mayor; es decir

$$R = x_{max} - x_{min}.$$

Sean  $R_1, R_2, \dots, R_m$  los rangos de las  $m$  muestras. El rango promedio es:

$$\bar{R} = \frac{R_1 + R_2 + R_3 + \dots + R_m}{m}.$$

Para la desviación típica  $\sigma$  se considera el estimador:

$$\hat{\sigma} = \frac{\bar{R}}{d_2},$$

donde, los valores de  $d_2$  se obtienen de la tabla Constantes para gráficos de control (Ver Apéndice B)

### 1.3.9. Fases de una carta de control

Existen dos fases para monitorear un proceso con una carta de control.

#### 1.3.9.1. Fase I

En esta fase se toma un conjunto de datos cuando se piensa que el proceso está bajo control, se estiman los parámetros del proceso, y se depura la información.

#### 1.3.9.2. Fase II

En esta Fase el proceso ya se encuentra estable y se lleva a cabo el monitoreo en línea del mismo (a través del tiempo).

# Capítulo 2

## Estado del arte

Durante la segunda guerra mundial la técnica estadística de gráficas de control univariadas fue la más utilizada a pesar de que los procesos y productos que se analizaban poseían en su gran mayoría dos o más características de calidad. Con el tiempo se dio la necesidad de aplicar herramientas estadísticas multivariadas para controlar en forma simultánea dos o más variables, [Herrera y Fontalvo \(2006\)](#). [Marroquín y Cantú \(2010\)](#), proponen la aplicación del control estadístico multivariante en un proceso de extrusión de película plástica, destacando la importancia del control estadístico multivariado.

Los procedimientos multivariados han venido cobrando cada día mayor atención, dado que los avances tecnológicos y las nuevas realidades facilitan y ponen en evidencia la necesidad de estudiar de manera simultánea más de una variable. De forma análoga a lo que ocurre con los procesos univariados, son muchos los casos en los que el interés se centra en la media del proceso, es decir, en el vector de medias. Sin embargo, dado que el monitoreo del vector de medias involucra el conocimiento de la matriz de covarianzas, se hace necesario estudiar esquemas de monitoreo de la matriz de covarianzas que sean eficientes. Para el monitoreo de la variabilidad en procesos multivariados, se ha usado tradicionalmente la carta de control de la varianza generalizada, propuesta por [Alt \(1985\)](#).

Existen actualmente alternativas para monitorear la variabilidad en procesos de control de calidad multivariados. Entre las opciones hay cartas tipo *Shewhart*, cartas tipo CUSUM y cartas tipo EWMA. Entre las cartas tipo *Shewhart* se encuentra la carta  $|S|$ , tal vez la más popular entre todas las cartas utilizadas para este propósito. Se construye de la siguiente manera ([Vargas \(2006\)](#)):

Suponiendo  $X_1, X_2, \dots, X_n$  una muestra aleatoria tomada de una distribución normal  $d$ -dimensional con parámetros  $\mu$  y  $\Sigma$ , para  $d = 2, 3, \dots$ , y  $n = 2, 3, \dots$ . Luego, el vector estadístico  $\bar{\mathbf{X}}$  de la muestra estadística y la matriz  $\mathbf{S}$  varianza-covarianza vienen dadas por:

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n X_i, \quad (2.1)$$

$$\mathbf{S} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{\mathbf{X}})(X_i - \bar{\mathbf{X}})', \quad (2.2)$$

de acuerdo con [Levinson \(2010\)](#), se puede demostrar que el determinante esperado y la línea central de la carta es:

$$\mathbb{E}(|S|) = b_1 |\Sigma|,$$

y la varianza del determinante está dada como,

$$\text{Var}(|S|) = b_2 |\Sigma|^2.$$

A partir de los esquemas para las cartas de control, los límites de control  $3\sigma$  tradicional (*Shewart*) son:

$$\begin{aligned} LCI &= |\Sigma| \left( b_1 - 3\sqrt{b_2} \right), \\ LC &= b_1 |\Sigma|, \\ LCS &= |\Sigma| \left( b_1 + 3\sqrt{b_2} \right). \end{aligned} \quad (2.3)$$

Esto, cuando se tiene que  $|\Sigma|$  es conocido (ver Anexo A: A.2).

En el caso en que  $|\Sigma|$  no sea conocida (ver Anexo A: A.3), se procede a realizar la respectiva estimación de dicho parámetro, por medio de la expresión

$$\frac{|S|}{b_1}, \quad (2.4)$$

el cual es un estimador insesgado de  $|\Sigma|$ . En este caso, los límites de control vendrían dados por:

$$\begin{aligned} LCI &= |S| \left( 1 - 3 \frac{\sqrt{b_2}}{b_1} \right), \\ LC &= |S|, \\ LCS &= |S| \left( 1 + 3 \frac{\sqrt{b_2}}{b_1} \right), \end{aligned} \tag{2.5}$$

donde, de manera general con  $d$ - dimensiones

$$\begin{aligned} b_1 &= \frac{1}{(n-1)^p} \prod_{i=1}^p (n-i), \\ b_2 &= \frac{1}{(n-1)^{2p}} \prod_{i=1}^p (n-i) \left[ \prod_{j=1}^p (n-j+2) - \prod_{j=1}^p (n-j) \right], \end{aligned} \tag{2.6}$$

donde  $n$  y  $p$ , representan el tamaño de muestra en cada subgrupopunto de muestreo y el número de variables que se tienen en cuenta en el proceso.

Para el caso particular en que  $p = 2$

$$b_1 = \frac{(n-2)}{(n-1)}; \quad b_2 = \frac{(n-2)(4n-2)}{(n-1)^3}. \tag{2.7}$$

**Nota:** En el caso en que el límite de control inferior resulte un valor menor que cero, este se toma como el valor mínimo que sería cero, ya que no tiene sentido que al momento de monitorear un proceso este contenga cantidades negativas.

[Montgomery y Wadsworth \(1972\)](#), utilizaron una aproximación asintótica normal para desarrollar la carta de control  $|S|$ . Esta carta fue construida usando los dos primeros momentos de  $|S|$ . Otro método sería utilizar la media y la varianza de  $|S|$  es decir,  $\mathbb{E}(|S|)$  y  $\sqrt{\text{Var}(|S|)}$ . [Alt y Smith \(1988\)](#) adoptaron un enfoque en la construcción de una carta de control para  $|S|$ , por el cual se estimó la verdadera varianza generalizada y su raíz cuadrada.

[Aparisi et al. \(1999\)](#) estudian las propiedades estadísticas de la carta de control multivariada (Carta  $|S|$ ), su trabajo se enfoca en la distribución de  $|S|$  y muestran como obtener límites de control adecuado. Anexando, que el poder de la tabla será estudiado y tienen como objetivo demostrar que solo depende de  $|\Sigma_1|/|\Sigma_0|$ , donde  $\Sigma_1$  es la matriz de covarianza fuera de control.

La carta de control de la varianza generalizada, tal y como fue presentada por [Ziegel \(2001\)](#) y [Jaupi \(2002\)](#), se construye usando el determinante del promedio de las matrices de covarianza como un estimador de la verdadera varianza generalizada y su raíz cuadrada. [Djauhari \(2005\)](#), hace notar que la forma estándar de usar el determinante del promedio de las matrices de covarianza de la muestra para la carta de control de variabilidad de proceso multivariados conduce a una estimación sesgada de los límites de control, debido a esto el autor presenta una carta de control mejorada cuyos límites de control son no sesgados y presentan una ARL más pequeña que la carta estándar y, por lo tanto, es más sensible a los cambios en la variabilidad del proceso.

[Olaya et al. \(2007\)](#), presentan una aplicación práctica de la teoría existente acerca del control estadístico de procesos para un proceso como el de extrusión de película plástica, cuya calidad es medida a través de la interacción de diferentes características.

[Morales y Vargas \(2008\)](#), proponen una carta EWMA para monitorear la variabilidad en procesos de control de calidad multivariados, haciendo uso de la varianza efectiva (que es la raíz  $p$ -ésima del determinante de la matriz de varianzas y covarianzas), en la cual se encuentran que esta carta resulta más eficiente que otras que usan como medida global de variabilidad a la varianza generalizada. Cuando hicieron los mayores cambios en la matriz de varianzas y covarianzas del proceso, la carta propuesta tuvo un desempeño similar al de las otras cartas.

[Hamed \(2014\)](#), propone una carta de control de calidad para la varianza generalizada  $|S|$  en la aplicación de las industrias fertilizante. El objetivo principal del autor es presentar una mejora en la carta de control para la varianza generalizada, teniendo en cuenta que una carta de control comúnmente requiere que se tomen muestras con un tamaño fijo a intervalos fijos. Finalmente, su aplicación muestra que la compañía debe usar cartas de control de calidad multivariante para determinar si el proceso no está bajo control porque la producción tiene varias variables correlacionadas, y el uso de cartas de control separados es engañoso porque las variables afectan conjuntamente el proceso. El uso de cartas de control univariados separados en una situación multivariada conduce a un error tipo I y la probabilidad de que un punto se trace correctamente en la carta no es igual a sus valores esperados.

A lo largo del tiempo, se han estudiado muchas características sobre esta carta, sin embargo no existe un estudio que muestre el efecto de la estimación de los parámetros en su desempeño y los pocos que se han hecho han sido con relación al vector de medias. Además, teniendo en cuenta la importancia de la Fase I en la construcción de las cartas de control y que muchas investigaciones se han centrado en determinar la importancia del tamaño de la muestra preliminar en el buen desempeño de la carta en Fase II. De forma similar, usando diferentes medidas de la eficiencia de la carta y para varios tipos de cartas [Chen \(1998\)](#), [Maravelakis et al. \(2002\)](#), [Willem y Wilbert \(2004\)](#), [Khoo y Lim \(2005\)](#), [Jensen et al. \(2006\)](#) y [Jarrett y Pan \(2009\)](#), concluyen por un lado que las cartas con límites estimados tienen un comportamiento diferente al de las cartas con límites conocidos y por el otro, que se debe aumentar el número de observaciones a usar en la Fase I para que el desempeño de las cartas sea más apropiado.

[Zhao y Driscoll \(2016\)](#), proponen evaluar el rendimiento bajo control de la carta  $c$  con un número promedio estimado de control de artículos no conformes, para estudiar el efecto de la estimación de parámetros en el desempeño de las cartas de control  $c$  con los límites de control ajustados de Bootstrap (resulta pertinente entonces abordar este tipo de estudios para el caso de la carta  $|S|$ ). Además, anotan que la integridad del monitoreo en Fase II depende de la confiabilidad de la estimación en Fase I. El desempeño de las cartas de control con parámetros estimados puede ser muy diferente del que tendrían si estos parámetros fueran conocidos.

[Peña y Álvarez \(2016\)](#) estudian la robustificación de la carta basada en la raíz cuadrada de la varianza muestral generalizada  $\sqrt{|S|}$  para el control de la variabilidad de un proceso normal bivariado en la Fase I de control, construida con observaciones sobre subgrupos racionales y utilizando los estimadores robustos MVE, MCD, estimador S. Como medida de desempeño se usaron el error cuadrático medio en la estimación de  $\Sigma_0$  y el sesgo absoluto en la estimación de  $\sqrt{|\Sigma_0|}$ , sobre los estimadores insesgados para cada uno de estos parámetros, respectivamente, construidos con los subgrupos racionales que quedan después del proceso de depuración realizado en la Fase I y que se consideran como el conjunto de datos que representa el estado de variación estable del proceso.

Por otra parte, [Yeh et al. \(2006\)](#) ofrecen una revisión exhaustiva de los gráficos de control multivariados entre 1990 y 2005 y diseñados para monitorear los cambios en una matriz de

covarianza al igual que [Ebadi et al. \(2020\)](#) que clasifican la investigación que ha aparecido previamente en la literatura, destacan las áreas desafiantes para la investigación y brindan algunas instrucciones para futuras investigaciones.

# Capítulo 3

## Estudio de simulación y resultados

En este capítulo estudiamos el comportamiento de la carta  $|S|$  en Fase I para dos escenarios. Consideramos procesos multivariados con  $p = 2$  características de calidad con vector de media ( $\mu = (0, 0)'$ ) y matrices de covarianza

$$\Sigma_{0_1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad \Sigma_{0_2} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix},$$

respectivamente. Para establecer el desempeño de la carta de control en Fase I para distintos tamaños de la muestra preliminar ( $n = 50, 10, 20, 30$  y  $100$ ) y números de muestras ( $m = 30, 50, 100, 200$  y  $1000$ ), se utilizarán procedimientos de simulación usando el software R, Versión 1.0.143.

### 3.1. Escenario 1

Mediante un estudio de simulación, considerando el vector  $\mu = (0, 0)$  y la matriz de covarianza  $\Sigma_{0_1}$  se lleva a cabo el estudio del efecto de la matriz de covarianza en el desempeño de la carta  $|S|$ .

#### 3.1.1. Resultados

Se asume, sin pérdida de generalidad, que

$$\mu_1 = (0, 0)' \quad \text{y} \quad \Sigma_{0_1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

A partir de este supuesto, se generan  $m = 30, 50, 100, 200, 1.000$  muestras de tamaño  $n \times p$ , estimando la matriz de covarianzas mediante la ecuación:

$$\bar{S} = \frac{1}{m} \sum_{i=1}^m S_i, \quad (3.1)$$

donde  $\bar{S}$  se define como el promedio de las  $S_i$  obtenidos en cada punto de muestreo. Luego se calculan los límites de control, por medio de la ecuación (3.2)

$$\begin{aligned} LCI &= |\bar{S}| \left( 1 - 3 \frac{\sqrt{b_2}}{b_1} \right), \\ LC &= |\bar{S}|, \\ LCS &= |\bar{S}| \left( 1 + 3 \frac{\sqrt{b_2}}{b_1} \right), \end{aligned} \quad (3.2)$$

donde,

$$b_1 = \frac{(n-2)}{(n-1)}; \quad b_2 = \frac{(n-2)(4n-2)}{(n-1)^3}, \quad (3.3)$$

luego,

- Se simulan 1000 replicas para cada combinación de tamaño  $n$  y  $m$ .
- A partir de cada muestra  $m_i$ , se determina la matriz de covarianza ( $S_i$ ) y se gráfica el determinante ( $|S_i|$ ) lo que nos permite hallar la proporción de falsas alarmas, dividiendo el número de veces que la carta da una señal fuera de control ( $|S_i| > LCS$ ) y el total de replicas. Este procedimiento se realizará para todos los escenarios que resulten de combinar los valores, con  $m = 30, 50, 100, 200, 1.000$  y  $n = 5, 10, 20, 30, 50, 100, 1.000$ .
- Una vez establecida la proporción de falsas alarmas para cada caso anterior, se ajustará un modelo que relacione estos valores con los valores de  $m$ , considerando  $m$  como variable independiente. Se espera que este modelo permita estimar la proporción de falsas alarmas reportadas por la carta en función del tamaño de la muestra

preliminar, considerando los diferentes valores de  $m$  (con una tasa de falsas alarmas nominal de 0.0027 (Gutiérrez y De la Vara (2004))).

En la Tabla 3.1 se muestran las proporciones de falsas alarmas para la carta  $|S|$  cuando  $\Sigma$  es estimada.

<b><math>K = 1000</math></b>	Tamaño de muestra ( $n$ )						
	Número de muestras ( $m$ )	5	10	20	30	50	100
30	0.006	0.015	0.003	0.006	0.018	0.005	0.004
50	0.001	0.006	0.001	0.014	0.007	0.008	0.003
100	0.004	0.028	0.008	0.008	0.003	0.007	0.001
200	0.025	0.007	0.005	0.005	0.009	0.004	0.001
1000	0.014	0.011	0.011	0.012	0.003	0.001	0.002

Tabla 3.1: Proporción de falsa alarma para la carta  $|S|$  cuando  $\Sigma$  es estimada, datos simulados bajo el Escenario 1.

De la Tabla 3.1 se puede observar que:

1. Las proporciones de falsas alarmas para la carta  $|S|$  son inicialmente altas cuando la matriz de varianza–covarianza es estimada a través de los  $m$  grupos, de tamaño  $n \times p$ . Por ejemplo, para las 1000 repeticiones, cuando  $n = 5$  y  $m = 30$ , la proporción de falsas alarmas es de 0.006 es relativamente alta comparada con la proporción de falsas alarma nominal, 0.0027. Mientras que cuando  $n = 1000$  y  $m = 1000$ , esta proporción de falsas alarmas es 0.002. Lo que hace notar, el efecto de la estimación de la matriz de covarianzas en el desempeño de la carta  $|S|$  cuando tanto el número de muestras como el tamaño de las mismas aumenta.
2. Por otra parte, observamos que a medida que el tamaño de las muestras ( $n$ ) y el número de muestras ( $m$ ) aumentan, la proporción de falsas alarmas tienden a disminuir para la mayoría de las combinaciones de  $m$  y  $n$ , y se aproximan a la tasa nominal, 0.0027.

### 3.1.2. Análisis gráfico

Las Figuras 3.1 y 3.2 muestran la proporción de falsas alarmas para cada combinación de  $n$  y  $m$ , considerando  $n$  variante en el primer caso y  $m$  en el segundo.

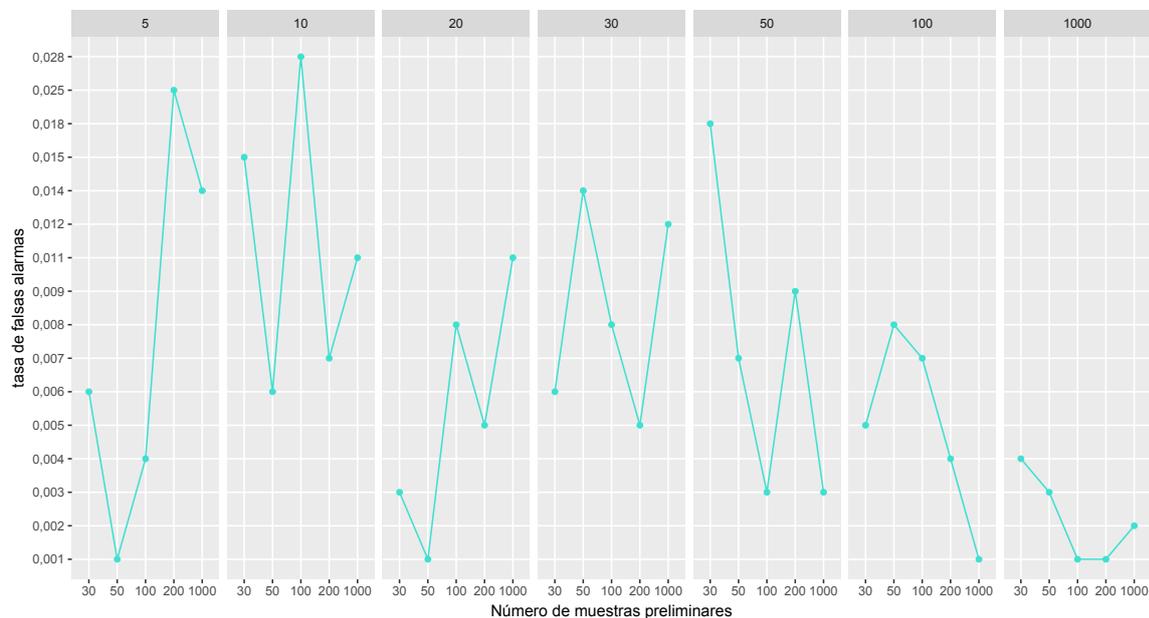


Figura 3.1: Resultados bajo el resultado 1, cuando el número de muestras ( $m$ ) es fijo y el tamaño de los subgrupos ( $n$ ) varían.

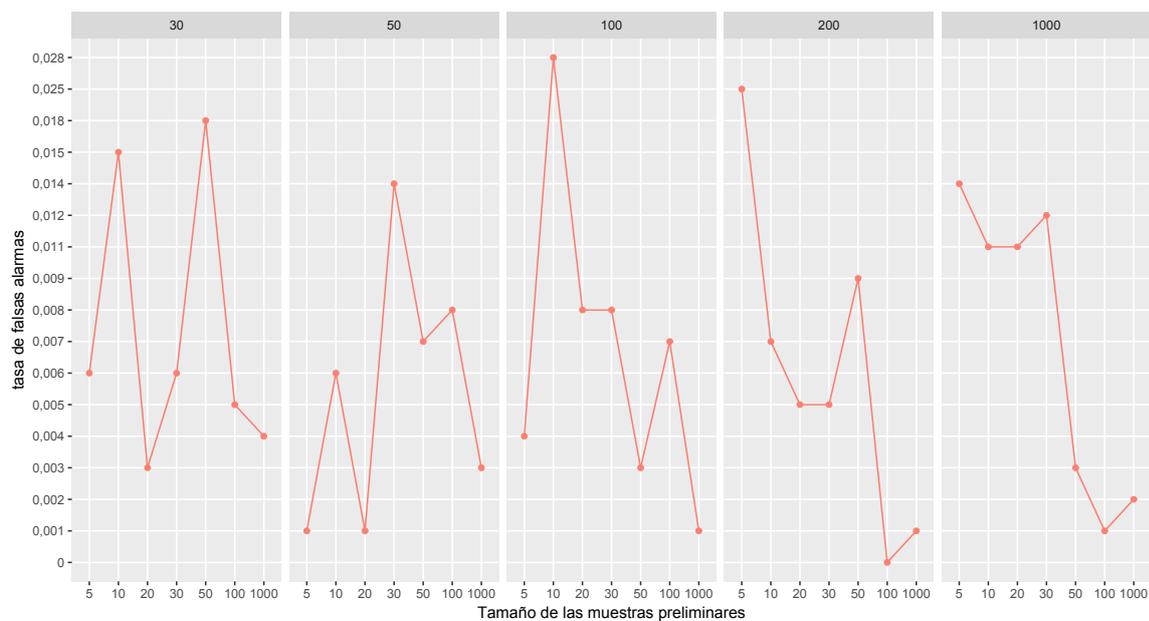


Figura 3.2: Resultados bajo el escenario 1, cuando el tamaño de los subgrupos ( $n$ ) es fijo y el número de muestras ( $m$ ) varían.

A partir de la Figura 3.1 para valores de  $m$  fijos pequeño existe una convergencia clara de los valores de la tasa de falsas alarmas hacia el valor nominal 0.0027 cuando  $m$  aumenta, mientras que en la Figura 3.2 cuando  $n$  tiene valores grandes no existe una clara convergencia del valor de la tasa de falsas alarmas hacia el valor nominal, al aumente  $m$ .

## 3.2. Escenario 2

Mediante un estudio de simulación, considerando el vector  $\mu = (0, 0)$  y la matriz de covarianza  $\Sigma_{0_2}$  se lleva a cabo el estudio del efecto de la matriz de covarianza en el desempeño de la carta  $|S|$ .

### 3.2.1. Resultados

Se asume sin pérdida de generalidad, que

$$\mu_2 = (0, 0)' \quad \text{y} \quad \Sigma_{0_2} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}.$$

A partir de este supuesto, se generan  $m = 30, 50, 100, 200$  y 1.000 muestras de tamaño  $n \times p$ , estimando la matriz de covarianzas mediante la ecuación (3.1). Luego se realiza el mismo proceso antes descrito bajo el escenario 1, considerando sigma estimada.

En la Tabla 3.2 se muestran las proporciones de falsas alarmas para la carta  $|S|$  cuando  $\Sigma$  es conocida.

<b><math>K = 1000</math></b>	Tamaño de muestra ( $n$ )						
	Número de muestras ( $m$ )	5	10	20	30	50	100
30	0.022	0.020	0.015	0.011	0.010	0.005	0.002
50	0.019	0.017	0.016	0.012	0.007	0.005	0.001
100	0.030	0.021	0.011	0.014	0.003	0.004	0.002
200	0.020	0.020	0.015	0.012	0.007	0.005	0.002
1000	0.024	0.015	0.008	0.015	0.008	0.005	0.002

Tabla 3.2: Proporción de falsa alarma para la carta  $|S|$  cuando  $\Sigma$  se estima, datos simulados bajo el Escenario 2.

De la Tabla 3.2 se puede observar que:

1. Las proporciones de falsas alarmas para la carta  $|S|$  son inicialmente altas cuando la matriz de varianza-covarianza es conocida a través de los  $m$  grupos, de tamaño  $n \times p$ . Por ejemplo, para las 1000 repeticiones, cuando  $n = 10$  y  $m = 30$ , la proporción de falsas alarmas es de 0.022 es relativamente alta comparada con la proporción de falsas alarma nominal, 0.0027. Mientras que cuando  $n = 1000$  y  $m = 1000$ , esta proporción de falsas alarmas es 0.002. Lo que hace notar, el efecto de la estimación de la matriz de covarianzas en el desempeño de la carta  $|S|$  cuando tanto el número de muestras como el tamaño de las mismas aumenta.
2. Por otra parte, observamos que a medida que el tamaño de las muestras ( $n$ ) y el número de muestras ( $m$ ) aumentan, la proporción de falsas alarmas tienden a disminuir para la mayoría de las combinaciones de  $m$  y  $n$  y se aproximan más similares a la tasa nominal, 0.0027.

### 3.2.2. Análisis gráfico

Las Figuras 3.3 y 3.4 muestran la proporción de falsas alarmas para cada combinación de  $n$  y  $m$ , considerando  $n$  variante en el primer caso y  $m$  en el segundo.

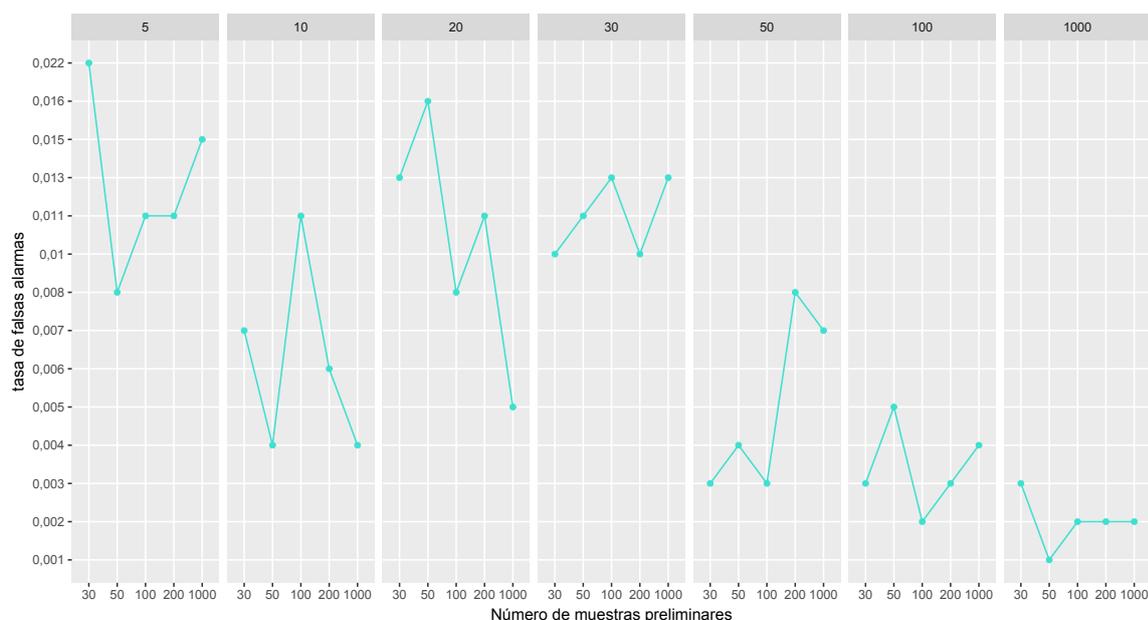


Figura 3.3: Resultados bajo el escenario 2, cuando el número de muestras ( $m$ ) es fijo y el tamaño de los subgrupos ( $n$ ) varían.

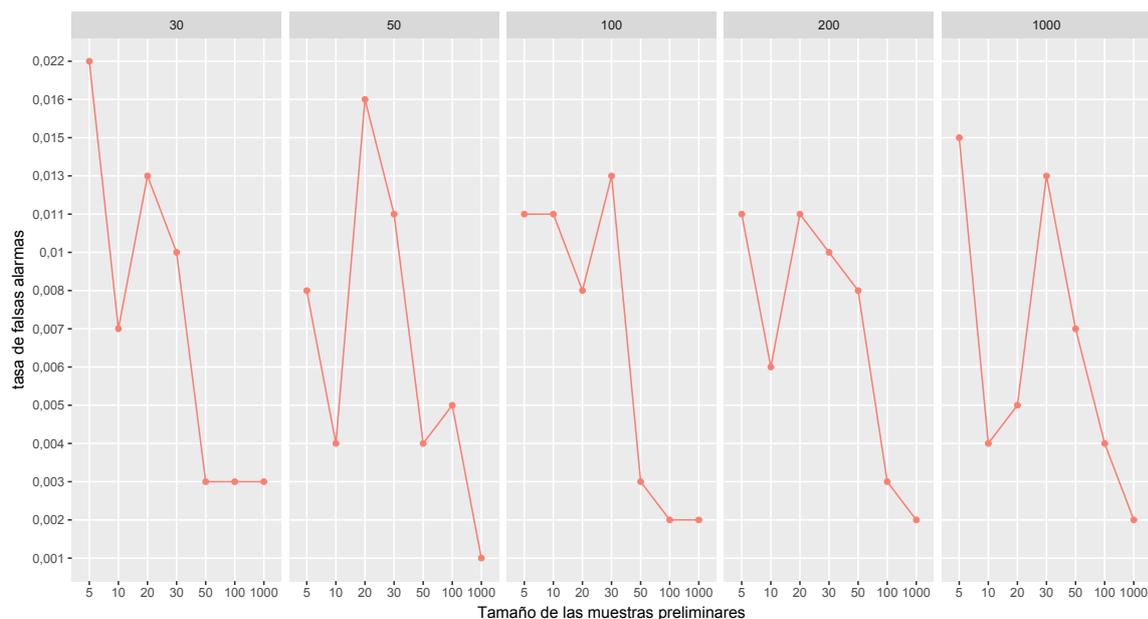


Figura 3.4: Resultados bajo el escenario 2, cuando el tamaño de los subgrupos ( $n$ ) es fijo y el número de muestras ( $m$ ) varían.

A partir de la Figura 3.3 para valores de  $m$  fijos pequeño existe una leve convergencia de los valores de la tasa de falsas alarmas hacia el valor nominal 0.0027 cuando  $n$  aumenta, mientras que en la Figura 3.4 cuando  $n$  tiene valores grandes existe una clara convergencia del valor de la tasa de falsas alarmas hacia el valor nominal, al aumente  $m$ .

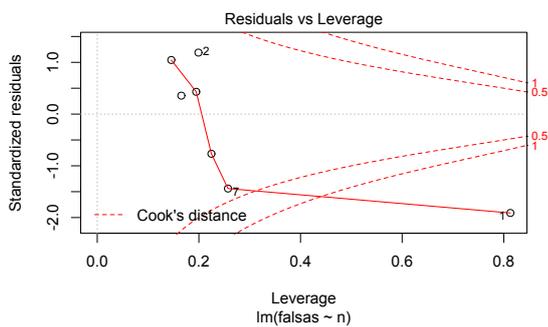
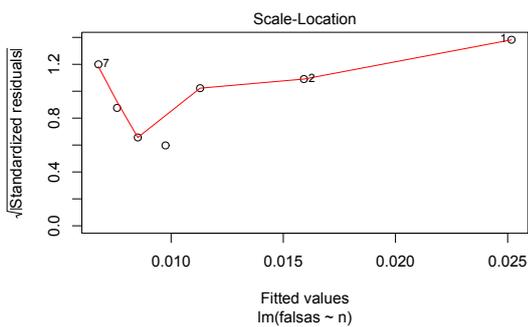
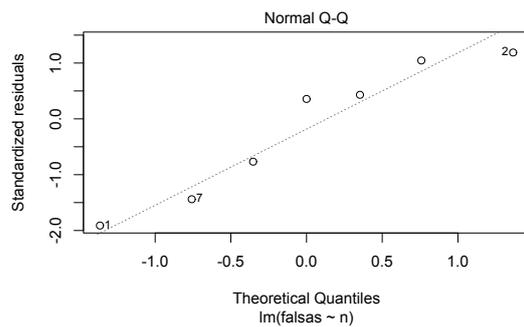
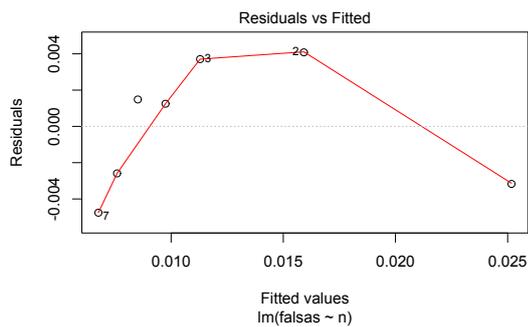
## Ajuste de modelos

Un modelo estimado para la tasa de falsas alarmas de la carta  $|S|$  usando  $n$  como la variable independiente en un modelo de regresión lineal simple, mientras se fija  $m$ , viene dado por:

En el caso cuando  $m = 30$ , el modelo de regresión ajustado es:

$$\hat{\alpha} = 0.006666 + 0.092521n \quad (3.4)$$

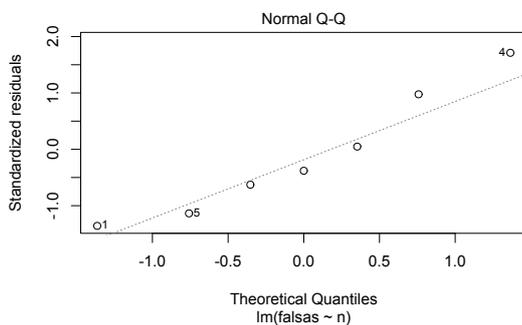
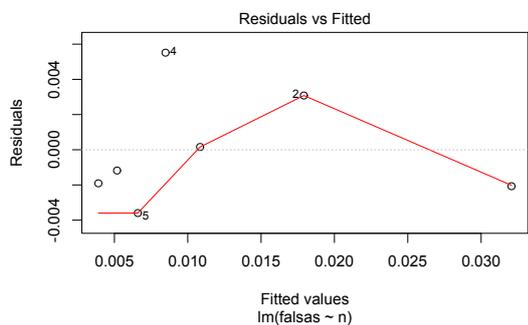
El coeficiente de determinación para el modelo de regresión en la ecuación (3.4) es  $R^2 = 0.75$ .

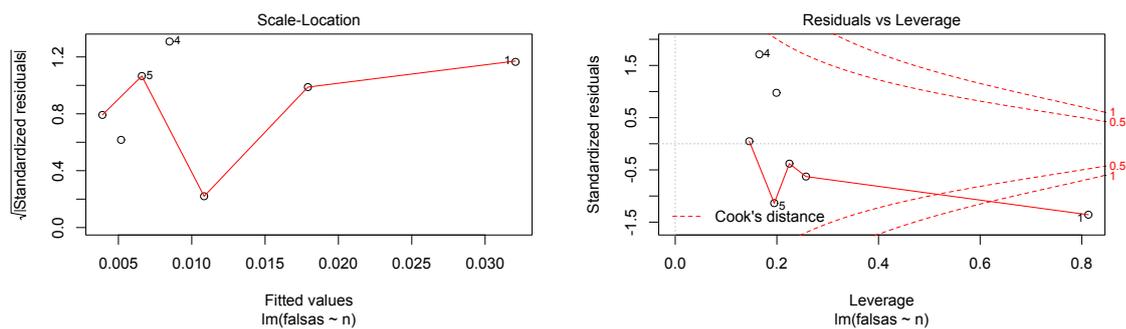


Para  $m = 100$ , el modelo de regresión ajustado es:

$$\hat{\alpha} = 0.003766 + 0.141528n. \tag{3.5}$$

El coeficiente de determinación para el modelo de regresión en la ecuación (3.5) es  $R^2 = 0.88$ .

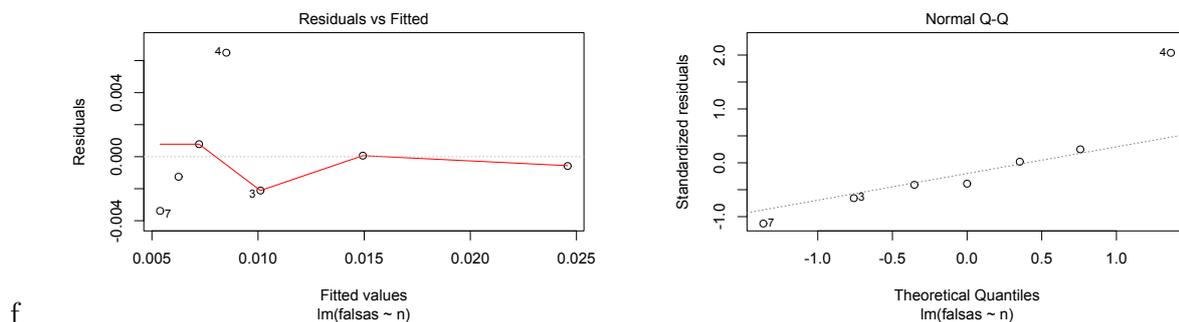




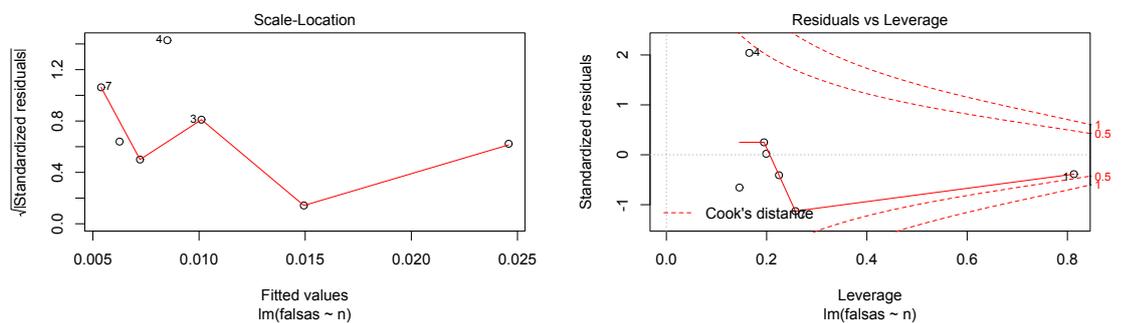
Para  $m = 1000$ , el modelo de regresión ajustado es:

$$\hat{\alpha} = 0.00529 + 0.09647n. \quad (3.6)$$

El coeficiente de determinación para el modelo de regresión en la ecuación (3.6) es  $R^2 = 0.78$ .



f



Cómo podemos ver en el gráfico 3.5, la relación inversamente proporcional entre el tamaño de la muestra  $n$  y la proporción de falsas alarmas, y estas a su vez se aproximan a la tasa de falsas alarma nominal. Los ajustes, según los cálculos de los  $R^2$ , es de 0.33, 0.20, y 0.08, respectivamente, lo cual quiere decir, que en los modelos las estimaciones no se

ajustan bien, considerando el hecho de que se trabaja con valores aleatorios a la variable real. Aunque técnicamente no sería correcto, podríamos decir algo así como que el modelo explica en un 33 %, 20 % y 8 % respectivamente, a la variable real; por lo que se realizó, la transformación del modelo lineal de la siguiente manera:

$$\hat{\alpha} = \beta_0 + \beta_1 Z; \quad \text{donde } Z = \frac{1}{n}.$$

Mejorando los ajustes de los modelos a los datos a 75 %, 88 % y 78 %, respectivamente. Estos modelos, ayudarían a obtener aproximaciones para la proporción de falsas alarmas de acuerdo con el tamaño de muestra preliminar ( $m$ ), y el tamaño de los subgrupos ( $n$ ) que se tengan. Además, se puede notar que los gráficos (a), (residuos versus valores ajustados) son diagramas de dispersión simple entre los residuos y los valores pronosticados. Deberían verse más o menos al azar. Esto es más o menos lo que vemos aquí. Los gráficos (b), (Q-Q normal) son gráficos de probabilidad normal. Dará una línea recta si los errores se distribuyen normalmente, pero se puede notar como en cada caso existen puntos que se desvían de la línea recta. Los gráficos (c), (Ubicación de escala), como el primero, debe parecer aleatorio, sin patrones. Por último, por medio de los gráficos (d), (la distancia de Cook) podemos saber qué puntos tienen la mayor influencia en la regresión (puntos de apalancamiento).

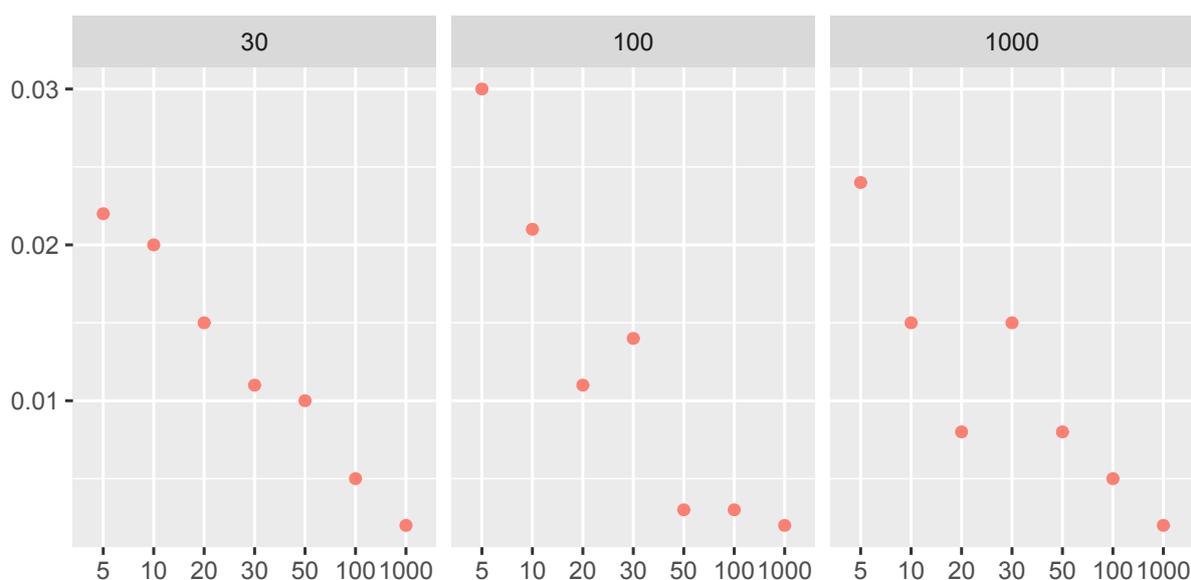


Figura 3.5: Diagrama de dispersión de los datos simulados.

# Capítulo 4

## Aplicación

La mayor parte de las investigaciones que se han realizado con relación a las cartas de control son de tipo univariado, donde se usa el promedio o la desviación estándar de una variable e inclusive el rango de esta. En algunos casos esta no es la solución más viable, por ejemplo, muchos procesos industriales son de naturaleza multivariada dado que la calidad de un producto depende de más de una variable. El control multivariado de procesos captura la relación en las variables asociadas al proceso, si se ignora esta correlación y se utilizan gráficos de control univariados para cada variable por separado se puede concluir erróneamente acerca del estado del proceso.

En variables continuas correlacionadas se han realizado muchas investigaciones, sin embargo se encuentra que se han realizado pocos trabajos que involucren el uso de cartas de control y estos usan esquemas univariados donde generalmente involucran el monitoreo del vector de medias, entre las variables estudiadas en el procesos.

En esta aplicación primeramente se muestran cartas control univariadas a manera de ilustración para detectar cambios significativos en el proceso, que permitan conocer el estado y el comportamiento de las características de calidad. Además se considera el uso de la carta de control para la varianza generalizada  $|S|$ , donde previamente se obtuvo que las variables aleatorias están altamente correlacionadas. Para efecto de mostrar el uso de las cartas de control multivariadas, consideramos los datos recopilados por el Servicio de Salud Ambiental Metropolitana de Chile correspondientes a las variables aleatorias: concentraciones de  $PM_{2.5}(X_1)$  y  $PM_{10}(X_2)$ , ambas medidas en  $\mu g/$  metro cúbico normalizado ( $Nm^3$ ). Estos datos están disponibles en [sinca.mma.gob.cl](http://sinca.mma.gob.cl) y se recopilaron en

2008 como valores promedio de 1 hora (por hora). Las concentraciones de  $PM_{2.5}$  y  $PM_{10}$  se observan en 13 estaciones de monitoreo ubicadas en diferentes zonas de la región Metropolitana de Chile. Estas estaciones se nombran como MS1 hasta MS13. Se considera la estación Padahuel (MS8) para nuestro análisis de datos, porque era la que tenía altos niveles de contaminación en el período de eventos críticos de la calidad del aire (del 1 de abril al 31 de agosto) en Santiago (Marchant et al. (2013)). Utilizamos datos de las concentraciones de  $PM_{2.5}$  y  $PM_{10}$  de la estación MS8 y los denominamos "PM2008MS8". Finalmente se muestra el efecto de la estimación de la matriz de covarianzas en el proceso. A partir de la distribución aproximada de los datos originales (se hizo una prueba de hipótesis, y se encontró que siguen aproximadamente una distribución normal bivariada), se simularon conjuntos de datos para diferentes valores de  $n$  y  $m$ , con los que posteriormente se ajustó un modelo que estudia la relación entre el tamaño de la muestra y la tasa de falsas alarmas, para un número de muestra fijo.

## 4.1. Análisis exploratorio de los datos

Inicialmente se realiza un análisis exploratorio de las variables  $PM_{2.5}$  y  $PM_{10}$  de nuestra base de datos para tener una idea de la forma que tienen los datos, encontrar patrones, identificar errores, plantear nuevas hipótesis o preguntas, identificar relaciones entre variables, inclusive empezar a encontrar respuestas a nuestras preguntas de investigación, además su posible distribución de probabilidad con sus parámetros de centralización; media, mediana y moda; así como sus respectivos parámetros de dispersión; varianza, desviación típica, coeficiente de variación, curtosis, coeficiente de asimétrica de Pearson y correlación entre las mismas. En la Tabla 4.1, se pueden ver los resultados obtenidos.

	$PM_{2.5}$	$PM_{10}$		$PM_{2.5}$	$PM_{10}$
Mínimo	1.00	1.00	Desv. est	50.29	95.01
Mediana	45.00	88.00	Coef. var	0.84	0.86
Media	59.62	110.7	Curtosis	3.83	4.50
Máximo	342.00	600.00	Asimetría	1.68	1.80
Varianza	2529.60	9026.75	Correlación	0.93	

Tabla 4.1: Análisis descriptivo de concentraciones de  $PM_{2.5}$  y  $PM_{10}$ .

A partir de la Tabla 4.1 se puede observar que: las concentraciones  $PM_{2.5}$  y  $PM_{10}$  en Pudahuel, para la variable  $PM_{2.5}$ , existe gran diferencia entre el máximo y el mínimo, se esperaría que el valor promedio fuera aproximadamente 171.5, pero fue 59.62, esto indica que los datos se encuentran dispersos. Mientras que para la variable  $PM_{10}$ , la diferencia es mucho mayor, de modo que los datos también se encuentran dispersos. La curtosis es positiva, para ambas variables, es decir, la mayoría de los datos están cerca de la media de manera que la forma de la curva o distribución será apuntada. Mediante el coeficiente de asimetría de Pearson, ambas variables siguen una distribución asimétrica positiva o a la derecha, así los valores se tienden a reunir más a la parte izquierda de la media. La fuerza y dirección de la relación lineal y proporcionalidad entre dos variables estadísticas es de 0.93, es decir están altamente correlacionadas. Ahora miremos el comportamiento de las variables gráficamente.

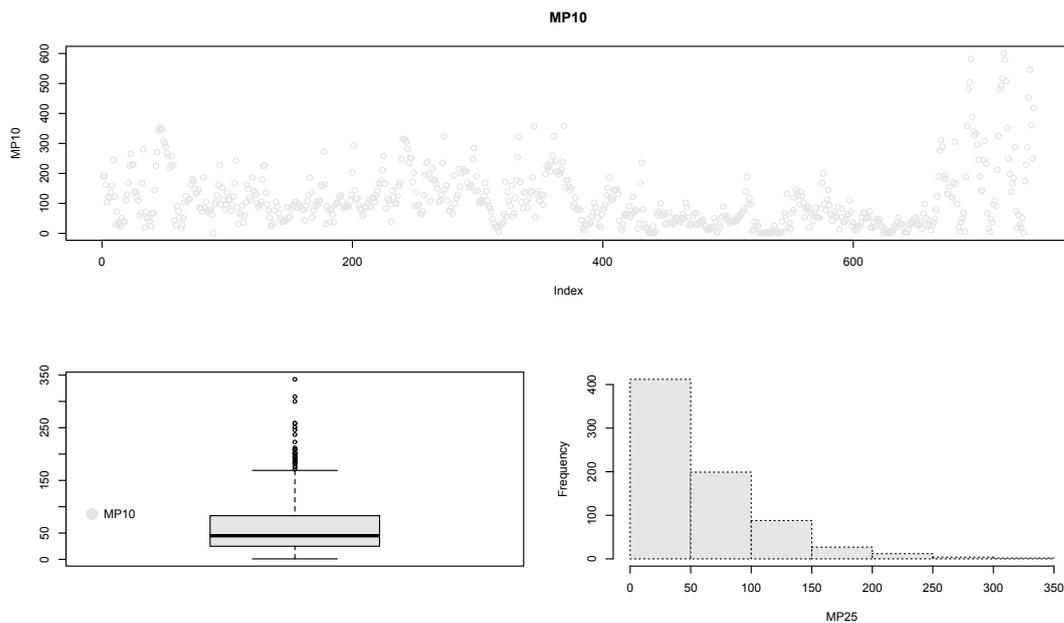


Figura 4.2: Diagrama de dispersión, diagrama de caja e histograma para  $PM_{10}$  con los datos de PM2008MS8.

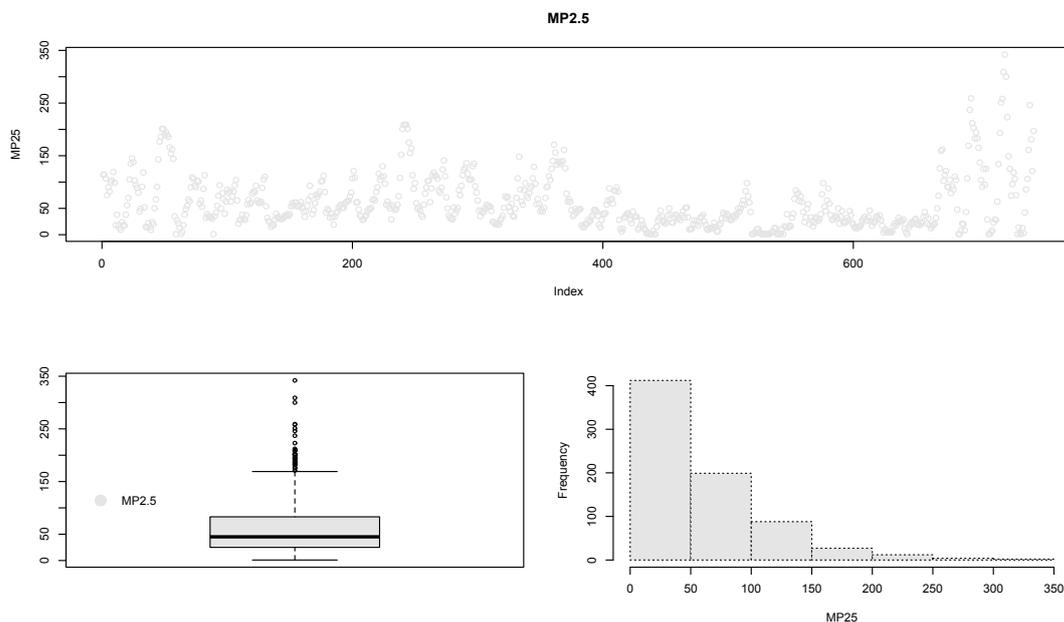


Figura 4.1: Diagrama de dispersión, diagrama de caja e histograma para  $PM_{2.5}$  con los datos de PM2008MS8.

En la Figura 4.1 notamos que no existe una tendencia marcada, al igual que en la Figura 4.2. Ahora, llevamos a cabo un análisis de correlación para detectar si las concentraciones de  $PM_{2.5}$  y  $PM_{10}$  de los datos de PM2008MS8 están estadísticamente asociadas. Además,

La Figura 4.3 muestra el diagrama de dispersión y el coeficiente de correlación de Pearson entre las concentraciones de  $PM_{2.5}$  y  $PM_{10}$ . A partir de esta figura, detectamos que existe una gran correlación (0.93) entre estas concentraciones en la estación MS8. Por lo tanto, para monitorear la contaminación ambiental urbana en Santiago, siguiendo la directriz del Ministerio de Salud de Chile, que indica que se deben considerar tanto  $PM_{2.5}$  como  $PM_{10}$ , proponemos utilizar una metodología basada en gráficos de control estadístico bivariado. Esta metodología nos permite modelar y monitorear las concentraciones de  $PM_{2.5}$  y  $PM_{10}$  simultáneamente, prediciendo adecuadamente los períodos críticos de contaminación.

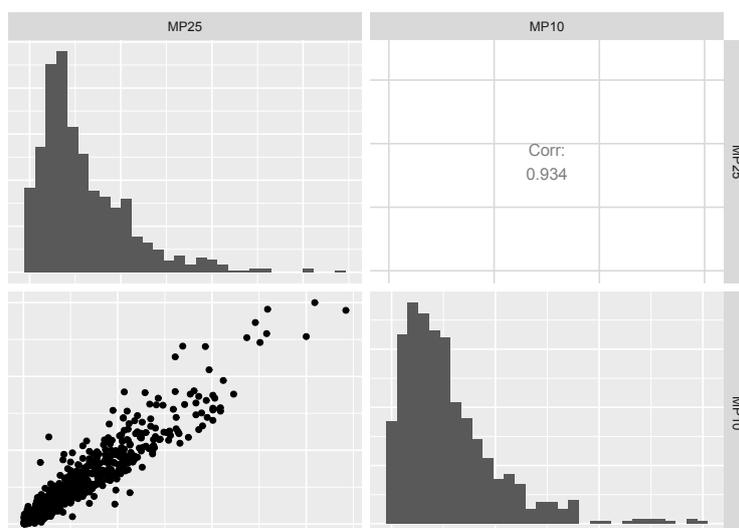


Figura 4.3: Matriz de gráfico de dispersión entre las concentraciones de  $PM_{2.5}$  y  $PM_{10}$  con los datos de PM2008MS8.

## 4.2. Cartas de control univariadas

En un proceso de producción cuando se monitorean variables continuas es importante la detección de los cambios en la variabilidad de las medidas de localización y variabilidad aplicadas. Por lo general, en el caso univariado se utilizan las cartas de control Shewhart, CUSUM y EWMA, que se construyen asumiendo normalidad e independencia en la medida de localización y de variabilidad. En caso de que este supuesto no se cumpla, se tendrá una considerable disminución en la sensibilidad de la carta de control para detectar señales de alarma. Notemos que la carta de control  $S$  tanto para la variable  $PM_{2.5}$  como para  $PM_{10}$ , no se encuentra bajo control, es decir, el proceso no es estable, siendo

este analizado de manera univariada, lo cual se puede evidenciar por medio de la Figura 4.4. Notemos que al existir asociación entre las variables es pertinente realizar la carta de control multivariada.

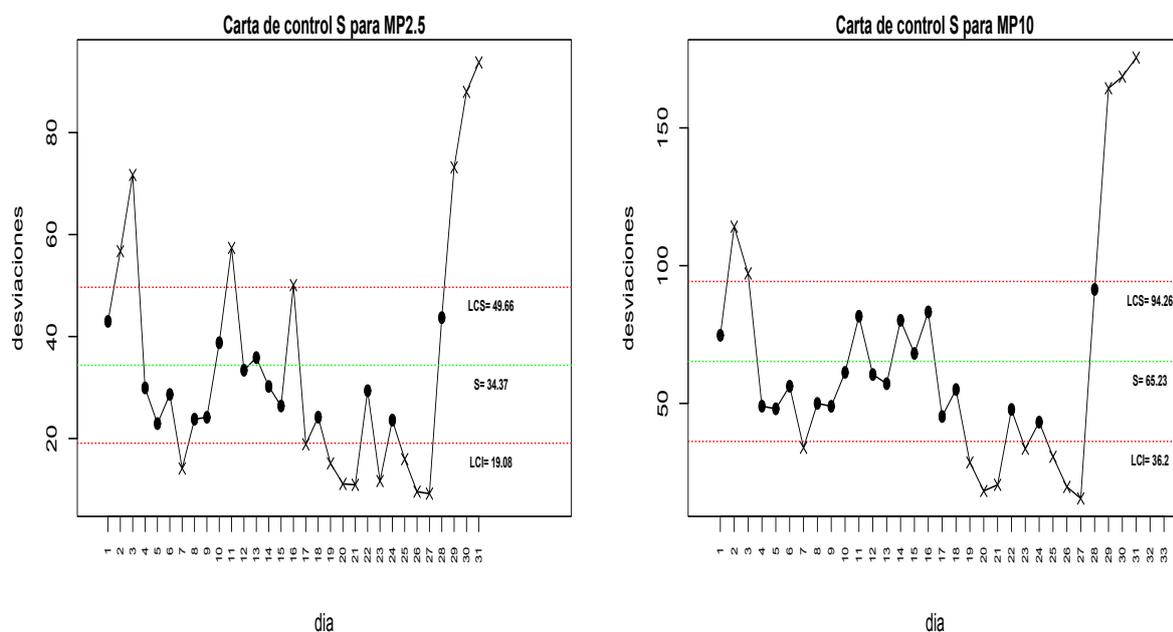


Figura 4.4: Cartas de Control S para  $PM_{2.5}$  y  $PM_{10}$ .

### 4.3. Carta de control $T^2$ Hotelling

La Figura 4.5 muestra el gráfico de control de Hotelling  $T^2$  para la estación MS8 que usan concentraciones de  $PM_{2.5}$  y  $PM_{10}$  recolectadas durante mayo–2008. Empleamos un paquete R llamado MSQC, relacionado con el control de calidad estadístico multivariado para construir este gráfico (Santos (2012)). Aquí, el número de subgrupos y el tamaño del subgrupo son  $m = 31$  días y  $n = 24$  horas, respectivamente. Teniendo en cuenta que la carta presentada en la Figura 4.4 indica que se controlan los niveles nocivos de  $PM_{2.5}$  y  $PM_{10}$ .

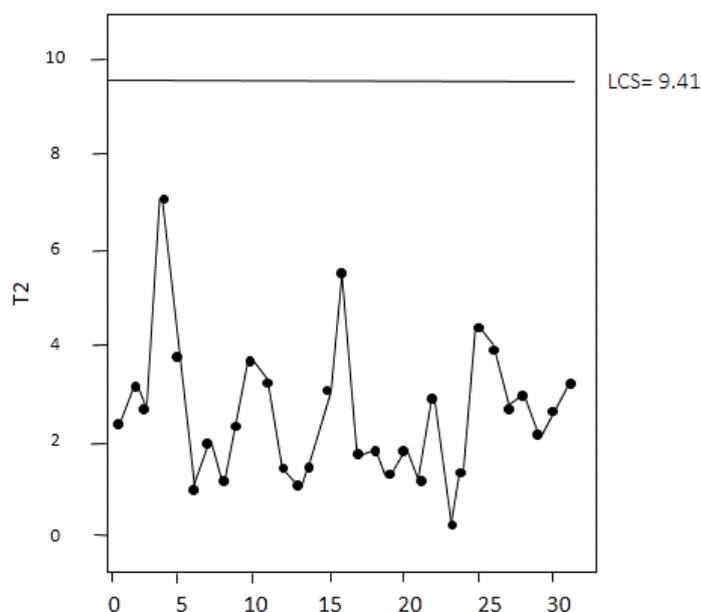


Figura 4.5: Carta de Control  $T^2$  de Hotelling para las concentraciones de  $PM_{2.5}$  y  $PM_{10}$  en la estación MS8.

#### 4.4. Carta de control $|S|$

Utilizando la ecuación (2.2), se calcularon las matrices de varianza-covarianza  $|S_i|$  para cada subgrupo  $i$  con sus respectivos determinantes  $S_i$ . En este caso, se consideró la  $\Sigma_0$  como desconocida, por lo que se utilizó el estimador definido para la carta. Luego, se utilizaron las ecuaciones (2.5) para encontrar los límites superior e inferior de la carta respectivamente, teniendo en cuenta los valores obtenidos para las constantes  $b_1$  y  $b_2$  con la ecuaciones (2.6). Finalmente, se graficaron los determinantes  $S_i$  con los límites calculados y se estudió el comportamiento de las observaciones de la muestra (Pombo et al. (2019))

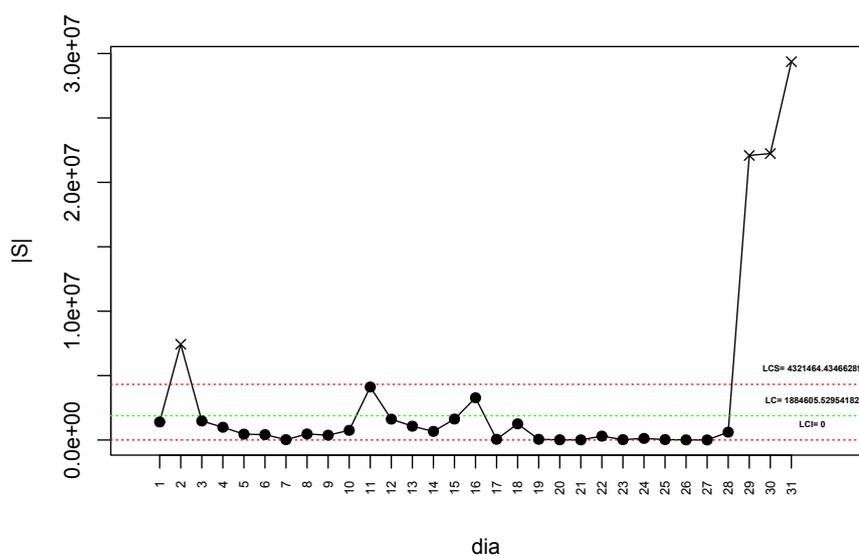


Figura 4.6: Carta de control multivariada  $|S|$  para  $MP_{2.5}$  y  $MP_{10}$ .

## Aplicación bajo simulación

En este caso, para el conjunto de datos  $PM2008MS8$  se encuentra el vector de medias ( $\hat{\mu}$ ) y la matriz de covarianzas ( $\hat{\Sigma}$ ), de donde se tiene que:

$$\hat{\mu} = (59.6, 110.7)' \quad \text{y} \quad \hat{\Sigma} = \begin{bmatrix} 866.6 & 1692.7 \\ 1692.7 & 3392.7 \end{bmatrix},$$

a partir de esta estimación se generan  $m$  muestras de tamaño  $n \times p$  de una normal bivariada con parámetros  $\hat{\mu}$  y  $\hat{\Sigma}$  ( $N_2 \sim (\hat{\mu}, \hat{\Sigma})$ ), con el objetivo de obtener  $\hat{\hat{\Sigma}}$  y posteriormente calcular los respectivos límites de control para la carta. Luego se simulan  $K = 1.000$  muestras de tamaño  $n \times p$ , donde se calculan las  $K$  matrices de covarianzas a graficar entre los límites antes calculados.

En la Tabla 4.2 se muestran las proporciones de falsas alarmas para la carta  $|S|$  usando valores reales. Simulando 1.000 observaciones en algunos de los casos en particular, se obtuvo que cuando  $n = 5$  y  $m = 30$ , la proporción de falsas alarmas es de 0.026, mientras que cuando  $n = 1000$  y  $m = 1000$ , la proporción de falsas alarma es 0.002.

$K = 1000$	Tamaño de muestra ( $n$ )	
Número de muestras ( $m$ )	5	1000
30	0.026	0.001
1000	0.026	0.002

Tabla 4.2: Proporción de falsa alarma para la carta  $|S|$ .

Con el objetivo de tener una estimación de los parámetros, por medio de la aplicación del enfoque bootstrap<sup>1</sup> la cual nos permite obtener estimaciones de medidas de precisión así como la realización de contrastes de hipótesis en aquellas situaciones en las que no se dispone de información acerca de las distribución muestral de un estadístico o en casos en los que la distribución muestral es dependiente de parámetros desconocidos (Efron (1982)), llegando a obtener:

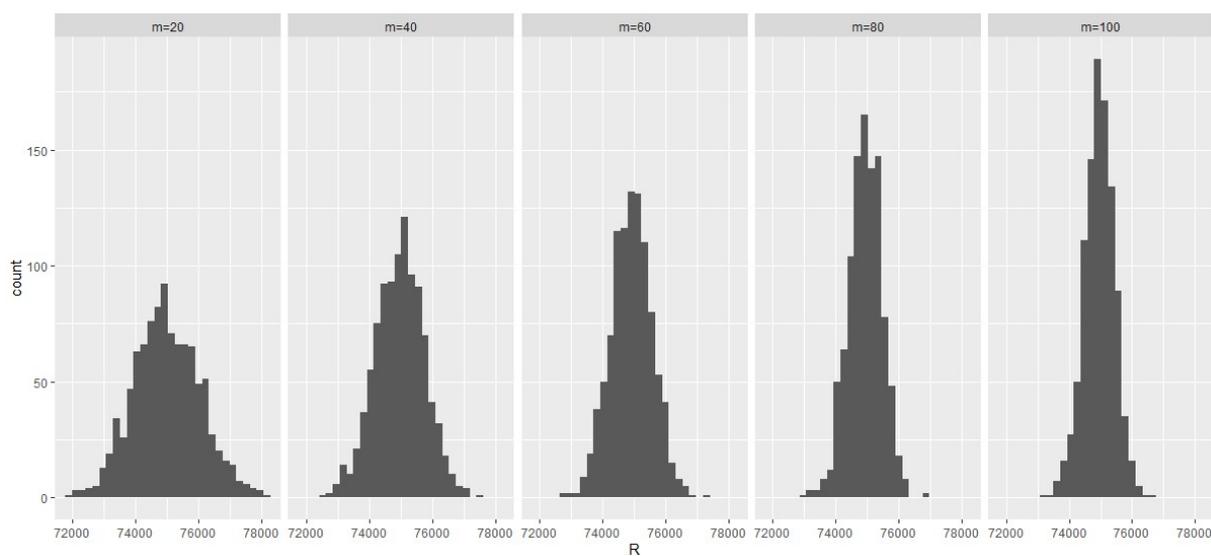


Figura 4.7: Distribución bootstrap para los resultados.

Por medio de la Figura 4.7, se implementa la técnica de re-muestreo *Bootstrap* para comprobar el grado de precisión de los resultados obtenidos bajo diferentes escenarios (20, 40, 60, 80 y 100), se puede notar que a medida que se aumenta el número de réplicas las estimaciones se hacen más precisas puesto que la variabilidad disminuye.

<sup>1</sup>es un método de remuestreo propuesto por Bradley Efron en 1979

# Capítulo 5

## Conclusiones y recomendaciones

### 5.1. Conclusiones

Con base en el análisis realizado y haciendo énfasis en los estudios de simulación y en el ejemplo aplicativo se puede concluir que:

1. La carta  $|S|$  depende significativamente tanto del tamaño de la muestra ( $n$ ) como el número de muestras ( $m$ ).
2. Valores pequeños de  $m$  y  $n$  implican una pobre estimación de la matriz de covarianza en control ( $\Sigma$ ), lo cual afecta el desempeño de la carta.
3. A medida que tanto  $n$  como  $m$  aumentan, las curvas correspondientes a la proporción de falsas alarmas para la carta  $|S|$  se aproximan a la tasa de proporción de falsas alarmas nominal, llegando a coincidir los perfiles en los diferentes escenarios.
4. No existe diferencia notoria en los resultados cuando la matriz de covarianza es  $\Sigma_{0_1}$  ó  $\Sigma_{0_2}$ , puesto que en ambos casos, la tasa de falsas alarmas se aproxima a la tasa de falsas alarma nominal.

$$\Sigma_{0_1} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad ; \quad \Sigma_{0_2} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

5. A partir de  $n = 1.000$  y  $m = 1.000$  aproximadamente, se puede decir que la carta  $|S|$  arroja una proporción de falsas alarmas cercana a 0.0027.

## 5.2. Recomendaciones para futuras investigaciones

Se presentan las siguientes recomendaciones a tener en cuenta para futuros trabajos de investigación en esta temática.

1. Ampliar el estudio, considerando datos generados a través de otros escenarios de simulación donde se mire el efecto del parámetro  $\mu$ , por ejemplo:

▪

$$\mu_1 = (2, 2)' \quad \text{y} \quad \Sigma_{0_1} = \begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix}$$

▪

$$\mu_2 = (1, 1)' \quad \text{y} \quad \Sigma_{0_2} = \begin{bmatrix} 10 & 1 \\ 1 & 10 \end{bmatrix}$$

2. Considerar un estudio donde el número de características de calidad aumente, por ejemplo  $p = 3$ , con su respectiva aplicación como ilustración.
3. A pesar de que las cartas tipo Shewhart multivariadas son tradicionalmente poco sensibles a cambios pequeños o moderados en la media de los datos ([Bersimis et al. \(2007\)](#)), seguramente el uso de una medida de tendencia central robusta en la carta de varianza efectiva  $|S|$ , mostrará una carta menos sensible a los cambios de los datos, en especial, cuando la muestra contiene valores extremos.
4. Ampliar el número de simulaciones para mayor certeza de los resultados, considerando el uso de un equipo adecuado, puesto que esta fue una de las limitaciones más visible al momento de llevar a cabo las respectivas compilaciones del mismo.
5. Explorar otros modelos que mejoren el ajuste entre la tasa de falsas alarmas ( $\alpha$ ) y el tamaño de la muestra ( $n$ ).

# Bibliografía

- Alt, F. (1985). Multivariate quality control. the encyclopedia of statistical sciences, kotz s., johnson, nl, read cr.
- Alt, F. B. y Smith, N. D. (1988). 17 multivariate process control. *Handbook of statistics*, 7:333–351.
- Aly, A. A., Mahmoud, M. A., y Hamed, R. (2016). The performance of the multivariate adaptive exponentially weighted moving average control chart with estimated parameters. *Quality and Reliability Engineering International*, 32(3):957–967.
- Aparisi, F., Jabaioyes, J., y Carrión, A. (1999). Statistical properties of the  $|s|$  multivariate control chart. *Communications in Statistics—theory and methods*, 28(11):2671–2686.
- Bergengruen, K. (2013). Causas comunes y causas especiales de la variación de las cartas de control.
- Bersimis, S., Psarakis, S., y Panaretos, J. (2007). Multivariate statistical process control charts: an overview. *Quality and Reliability engineering international*, 23(5):517–543.
- Blanco, E. E. (2012). Cartas  $r$  y  $s$  con límites de control estimados. *Departamento de Estadística*.
- Carro, R. y González Gómez, D. A. (2012). Control estadístico de procesos.
- Chakraborti, S. y Human, S. W. (2006). Parameter estimation and performance of the  $p$ -chart for attributes data. *IEEE Transactions on Reliability*, 55(3):559–566.
- Chen, G. (1998). The run length distributions of the  $r$ ,  $s$  and  $s^2$  control charts when is estimated. *Canadian Journal of Statistics*, 26(2):311–322.

- Chen, Y., Birch, J. B., y Woodall, W. H. (2016). Effect of phase i estimation on phase ii control chart performance with profile data. *Quality and Reliability Engineering International*, 32(1):79–87.
- Coronado, R. (2007). Control estadístico en el proceso de fabricación de equipos de refrigeración comercial.
- Crosby, P. B. (1996). *Los principios absolutos del liderazgo*. Number 658.4/C94aE.
- Deming, E. (1982). *Calidad, productividad y Competitividad. Fuera de la Crisis*. México: Editorial Díaz de Santos SA.
- Djauhari, M. A. (2005). Improved monitoring of multivariate process variability. *Journal of Quality Technology*, 37(1):32–39.
- Ebadi, M., Chenouri, S., y Lin, D. K. (2020). Statistical monitoring of covariance matrix in multivariate processes: A literature review. *arXiv preprint arXiv:2002.06159*.
- Efron, B. (1982). *The jackknife, the bootstrap and othe resampling plans*. SIAM.
- Feigenbaum, A. V. (1991). Total quality control. *tqc*.
- García, S. L. (2014). *Control multivariante estadístico de variables discretas tipo poisson*. PhD thesis.
- Gutiérrez, Humberto, f. y De la Vara, S. R. (2004). *Control estadístico de calidad y seis sigma*. MCGRALL-HILL.
- Hamed, M. (2014). Generalized variance chart for multivariate quality control process procedure with application. *Applied Mathematical Sciences*, 8(163):8137–8151.
- Hammurabi, R. d. B. (1970). *código de Hammurabi*. Instituto Técnico de Materiales y Construcciones.
- Harrington, H., Rosas, G., et al., (1993). *Mejoramiento de los procesos de la empresa*.
- Herrera, R. y Fontalvo, T. (2006). Seis sigma métodos estadísticos y sus aplicaciones. *Ecoe Ediciones*, page 56.
- Hervás, J. A. (2015). Manual de control estadístico para la gestión de procesos.

- Ishikawa, K. (1943). Diagrama causa-efecto. *Recuperado el*, 15.
- Jarrett, J. E. y Pan, X. (2009). Multivariate process control charts and their use in monitoring output quality: a perspective. *International Journal of Industrial and Systems Engineering*, 4(5):471–482.
- Jaupi, L. (2002). *Contrôle de la qualité: MSP, analyse des performances et contrôle de réception*. Dunod.
- Jensen, W. A., Jones-Farmer, L. A., Champ, C. W., y Woodall, W. H. (2006). Effects of parameter estimation on control chart properties: a literature review. *Journal of Quality Technology*, 38(4):349–364.
- Juran, J. M. (2001). Manual de control de calidad - 5 edición 2 tomos.
- Juran, J. M., Gryna, F. M., y Bingham, R. S. (1983). *Manual de Control de la Calidad*. Reverté.
- Khoo, M. B. y Lim, E. (2005). An improved r (range) control chart for monitoring the process variance. *Quality and Reliability Engineering International*, 21(1):43–50.
- Levinson, W. A. (2010). *Statistical process control for real-world applications*. CRC Press.
- Mahmoud, M. y Woodall, W. (2004). Phase i analysis of linear profiles with calibration applications. *Technometrics*, 46(4):380–391.
- Maravelakis, P., Panaretos, J., y Psarakis, S. (2002). Effect of estimation of the process parameters on the control limits of the univariate control charts for process dispersion. *Communications in Statistics-Simulation and Computation*, 31(3):443–461.
- Marchant, C., Leiva, V., Cavieres, M. F., y Sanhueza, A. (2013). Air contaminant statistical distributions with application to pm10 in santiago, chile. In *Reviews of Environmental Contamination and Toxicology Volume 223*, pages 1–31. Springer.
- Marroquín, E. y Cantú, M. (2010). Una gráfica de control combinada para identificar señales fuera de control en procesos multivariados. *Ingeniería, investigación y tecnología*, 11(4):453–460.

- Montgomery, D. y Wadsworth, H. (1972). Some techniques for multivariate quality control applications. In *ASQC Technical Conference Transactions*, volume 26, pages 427–435.
- Montgomery, D. C. (2007). *Introduction to statistical quality control*. John Wiley & Sons.
- Morales, V. H. y Vargas, J. A. (2008). Carta ewma con varianza efectiva para monitorear variabilidad en procesos de control de calidad multivariados. *Revista Colombiana de Estadística*, 31(1):131–143.
- Neria, G. (2014). Control estadístico de la calidad.
- Ojeda, E. (2015). Estadística y control de calidad.
- Olaya, J., Restrepo, J. M., y Escobar, R. (2007). Aplicación del control estadístico multivariante en un proceso de extrusión de película plástica. *Scientia et Technica*, 5(37):333–338.
- Peña, E. D. y Álvarez, N. G. G. (2016). Robustificación de la carta de control multivariada  $\sqrt{|S|}$  en la fase i de control. *Revista de la Facultad de Ciencias*, 5(2):12–37.
- Périsse, M. C. (2007). Control de la calidad utilizando excel.
- Plaza, A. (2018). Diseño de gráficos de control multivariantes basado en la combinación lineal de variables normales. B.S. thesis, Espol.
- Pombo, R. J. B., Guerrero, A. P. A., y Acosta, R. J. H. (2019). Propuesta de carta de control multivariada utilizando la media winsorizada basada en la carta  $|s|$ /proposed multivariate control chart using winsorized mean based on the  $|s|$  chart. *Revista Ingeniería*, 29(1):65–79.
- Psarakis, S., Vyniou, A. K., y Castagliola, P. (2014). Some recent developments on the effects of parameter estimation on control charts. *Quality and Reliability Engineering International*, 30(8):1113–1129.
- Raubenheimer, L. y Van der Merwe, A. J. (2015). Bayesian control chart for nonconformities. *Quality and Reliability Engineering International*, 31(8):1359–1366.
- Rendón, H. (2013). Control estadístico de calidad. *Medellín: Centro Editorial de la Facultad de Minas Universidad Nacional de Colombia*.

- Rivera, G. (2011). Cartas de control para datos funcionales. *México. Centro de investigación en matemáticas, AC Tesis de maestría.*
- Rojas, A. R.-F. (2006). Control estadístico de procesos. *Universidad Pontificia, Madrid.*
- Ruiz, D., Sánchez, I., Sánchez, B., Santiago, M., Vazquez, D., Velasco, S., y Vidal, L. (2016). Herramientas básicas para el control de calidad.
- Sales, M. (2002). Diagrama de pareto.
- Santos, E. (2012). *Multivariate statistical quality control using R*, volume 14. Springer Science & Business Media.
- Shewhart, W. (1924). Some applications of statistical methods.
- SPC, C. G. (2013). Introducción al control estadístico de procesos.
- Testik, M. C. (2007). Conditional and marginal performance of the poisson cusum control chart with parameter estimation. *International Journal of Production Research*, 45(23):5621–5638.
- Vargas, J. (2006). Control estadístico de calidad. *Universidad Nacional de Colombia, Bogotá, Colombia.*
- Willem, A. y Wilbert, K. (2004). Estimation in shewhart control charts: effects and corrections. *Metrika*, 59(3):207–234.
- Yeh, A. B., Lin, D. K., y McGrath, R. N. (2006). Multivariate control charts for monitoring covariance matrix: a review. *Quality Technology & Quantitative Management*, 3(4):415–436.
- Zhao, M. J. y Driscoll, A. R. (2016). The c-chart with bootstrap adjusted control limits to improve conditional performance. *Quality and Reliability Engineering International*, 32(8):2871–2881.
- Ziegel, E. R. (2001). Introduction to statistical quality control. *Technometrics*, 43(4):497.

# Apéndice A

## Demostraciones

En esta sección se presentan los diferentes anexos necesarios para completar el desarrollo de este trabajo. En el anexo A se llevan a cabo demostraciones relacionadas con los límites de control de los gráficos.

1. Límites de control para el gráfico de control  $S$ .

$$\begin{aligned} LCI &= \mathbb{E}(|S|) - 3\sqrt{\text{Var}(|S|)} \\ &= \bar{S} - 3\frac{\bar{S}}{C_4}\sqrt{1 - C_4^2} \\ &= \left(1 - \frac{3}{C_4}\sqrt{1 - C_4^2}\right)\bar{S} \\ &= B_3\bar{S} \end{aligned}$$

$$\begin{aligned} LC &= \mathbb{E}(|S|) \\ &= \bar{S} \end{aligned} \tag{A.1}$$

$$\begin{aligned} LCS &= \mathbb{E}(|S|) + 3\sqrt{\text{Var}(|S|)} \\ &= \bar{S} + 3\frac{\bar{S}}{C_4}\sqrt{1 - C_4^2} \\ &= \left(1 + \frac{3}{C_4}\sqrt{1 - C_4^2}\right)\bar{S} \\ &= B_4\bar{S} \end{aligned}$$

2. Límites de control para la carta de control  $|S|$  cuando  $\Sigma$  es conocida.

$$\begin{aligned}
LCI &= \mathbb{E}(|S|) - 3\sqrt{\text{Var}(|S|)} \\
&= b_1|\Sigma| - 3\sqrt{b_2|\Sigma|^2} \\
&= b_1|\Sigma| - 3|\Sigma|\sqrt{b_2} \\
&= |\Sigma|(b_1 - 3\sqrt{b_2}) \\
LC &= \mathbb{E}(|S|) \\
&= b_1|\Sigma| \\
LCS &= \mathbb{E}(|S|) + 3\sqrt{\text{Var}(|S|)} \\
&= b_1|\Sigma| + 3\sqrt{b_2|\Sigma|^2} \\
&= b_1|\Sigma| + 3|\Sigma|\sqrt{b_2} \\
&= |\Sigma|(b_1 + 3\sqrt{b_2})
\end{aligned} \tag{A.2}$$

3. Límites de control para la carta de control  $|S|$  cuando  $\Sigma$  se estima.

$$\begin{aligned}
LCI &= \mathbb{E}(|S|) - 3\sqrt{\text{Var}(|S|)} \\
&= b_1|\Sigma| - 3\sqrt{b_2|\Sigma|^2} \\
&= b_1\frac{|S|}{b_1} - 3\frac{|S|}{b_1}\sqrt{b_2} \\
&= |S| - 3|S|\frac{\sqrt{b_2}}{b_1} \\
&= |S|\left(1 - 3\frac{\sqrt{b_2}}{b_1}\right) \\
LC &= \mathbb{E}(|S|) \\
&= b_1|\Sigma| \\
&= b_1\frac{|S|}{b_1} \\
&= |S| \\
LCS &= \mathbb{E}(|S|) + 3\sqrt{\text{Var}(|S|)} \\
&= b_1|\Sigma| + 3\sqrt{b_2|\Sigma|^2} \\
&= b_1\frac{|S|}{b_1} + 3\frac{|S|}{b_1}\sqrt{b_2} \\
&= |S| + 3|S|\frac{\sqrt{b_2}}{b_1} \\
&= |S|\left(1 + 3\frac{\sqrt{b_2}}{b_1}\right)
\end{aligned} \tag{A.3}$$

# Apéndice B

## Constantes para Gráficos de Control

En este anexo, se incluye una tabla de constantes necesarias para la realización de los diferentes gráficos de control clásicos. La siguiente tabla, tiene relacionada las constantes para gráficos de control, antes aplicadas en la teoría.

n	$B_3$	$B_4$	$c_4$	$d_2$	n	$B_3$	$B_4$	$c_4$	$d_2$
2	0.000	3.267	0.798	1.128	14	0.406	1.594	0.981	3.407
3	0.000	2.568	0.886	1.693	15	0.428	1.572	0.982	3.472
4	0.000	2.266	0.921	2.059	16	0.448	1.552	0.983	3.532
5	0.000	2.089	0.940	2.326	17	0.466	1.534	0.985	3.588
6	0.030	1.970	0.952	2.534	18	0.482	1.518	0.985	3.640
7	0.118	1.882	0.959	2.704	19	0.497	1.503	0.986	3.689
8	0.185	1.815	0.965	2.847	20	0.510	1.490	0.987	3.735
9	0.239	1.761	0.969	2.970	21	0.523	1.477	0.988	3.778
10	0.284	1.716	0.973	3.078	22	0.534	1.466	0.988	3.819
11	0.321	1.679	0.975	3.173	23	0.545	1.455	0.989	3.858
12	0.354	1.646	0.978	3.258	24	0.555	1.445	0.989	3.895
13	0.382	1.618	0.979	3.336	25	0.565	1.435	0.990	3.931

Tabla B.1: Constante para gráfico de control.

# Apéndice C

## Código de simulación en R

En este anexo se da a conocer el código en R (R Development Core Team 2012) utilizado para la realización de los diferentes gráficos de control y las simulaciones bajo los dos escenarios estudiados.

```
#-----  
#Paquetes necesarios  
#-----  
library(e1071); library(GGally); library(ggthemes)  
library(ggplot2); library(gmodels); library(Hmisc)  
library(MASS); library(moments); library(MSQC)  
library(mvtnorm); library(readxl); library(qcr)  
#-----  
##Simulación (Capitulo 3)  
#-----  
#Simulación para resultados de la tabla 3.1 y 3.2  
#-----  
#Cargamos los paquetes  
#Se crea una función  
Resultado<- function(n,m,K)  
  #p características de calidad  
  p = 2  
  #mu fijo
```

```

mu = c(0,0)
#sigma conocida
sigma = matrix(c(2,1,1,2),2)
#matrices vacias donde se guardarán resultados
sigmaest<-c() ;
datos<-c() ;
Resultado1<-c() ;
Resultado2<-c() ;
#Generación de datos
.
.
.
#Constantes para el calculo de los limites de control
i<- c(1,2)
j<- c(1,2);
dS<- det(SigmaEstimadoInicial);
b1<- round(((1/(n-1)^2)* prod(n-i)),3);
b2<- round((((1/(n-1)^2*p)*(prod(n-i)))*((prod(n-j+2) - prod(n-j)))),3);
#Calculo de limites de control
LCS1<- round((det(sigma))*(b1+(3*(sqrt(b2))))),4);
LC1<- round(det(sigma),4);
LCI1<- round((det(sigma))*(b1-(3*(sqrt(b2))))),4);
#Condicion cuando el LCI es menor que cero.
if(LCI1<0) LCI2<-0
else
  LCI1<- LCI1
#Calculo de limites de control
LCS2<- round((dS/b1)*(b1+(3*(sqrt(b2))))),4);
LC2<- round(dS,4);
LCI2<- round((dS/b1)*(b1-(3*(sqrt(b2))))),4);
#Condicion cuando el LCI es menor que cero.
if(LCI1<0) LCI1<-0

```

```

else
  LCI1<- LCI1
#Generación de datos a graficar
sigmaestK<-c()
datos<-c()
Resultado1<-c()
Resultado2<-c()
for(i in 1:K)
  datos= replicate(K,list((mvrnorm(n,mu,sigma))))
  sigmaestK[i]<-list(cov(datos[[i]]))

for(j in 1:K)
  Resultado1[j]<- c(det(sigmaestK[[j]]))
  Resultado2[j]<- c(det(sigmaestK[[j]]))

#Luego se generan contadores RL y RL2 para tener la proporción de
#falsas alarmas para la carta |S|
data.frame(RL,RL/K,RL2,RL2/K)

#-----
#Gráficos 3.3 y 3.4.
#-----
#Escenario 1
#-----
# Datos a utilizar bajo el escenario 1
# Cargar el directorio setwd("F:/Tesis Códigos R Studio")
simula_bb<-read.table("Escenario1.txt",header=TRUE)
simula_bb$n<-factor(simula_bb$n)
simula_bb$m<-factor(simula_bb$m)
simula_bb$sigma<-factor(simula_bb$sigma)
pg <- ggplot(simula_bb, aes(m,Resultados,color=sigma))+
  geom_point()+

```

```

geom_line(group=simula_bb$sigma)+
geom_smooth(method = "loess", se = FALSE)+
facet_grid(~n) +
labs(x="Número de muestras preliminares",y="Número de falsas alarmas",
      #title="Escenario 1",subtitle="m: fijo, n: variable",
      color =c(expression(Sigma)))
print(pg)
#-----
pg1 <- ggplot(simula_bb, aes(n,Resultados,color=sigma))+
  geom_point()+
  geom_line(group=simula_bb$sigma)+
  geom_smooth(method = "loess", se = FALSE)+
  facet_grid(~m) +
  labs(x="Tamaño de las muestras preliminares",y="Número de falsas alarmas",
        #title="Escenario 1",subtitle="m: variable, n: fijo",
        color =c(expression(Sigma)))
print(pg1)
#-----
#Escenario 2
#-----
simula_bb<-read.table("Escenario2.txt",header=TRUE)
simula_bb$n<-factor(simula_bb$n)
simula_bb$m<-factor(simula_bb$m)
simula_bb$sigma<-factor(simula_bb$sigma)

pg3 <- ggplot(simula_bb, aes(m,Resultados,color=sigma))+
  geom_point()+
  geom_line(group=simula_bb$sigma)+
  geom_smooth(method = "loess", se = FALSE)+
  facet_grid(~n) +
  labs(x="Número de muestras preliminares",y="Número de falsas alarmas",
        #title="Escenario 2",subtitle="m: fijo, n: variable",

```

```

        color =c(expression(Sigma)))
print(pg3)
#-----
pg4 <- ggplot(simula_bb, aes(n,Resultados,color=sigma))+
  geom_point()+
  geom_line(group=simula_bb$sigma)+
  geom_smooth(method = "loess", se = FALSE)+
  facet_grid(~m) +
  labs(x="Tamaño de las muestras preliminares",y="Número de falsas alarmas",
       #title="Escenario 2",subtitle="m: variable, n: fijo",
       color =c(expression(Sigma)))
print(pg4)
#-----
#Gráficos de dispersión (Figura 3.5)
#-----
#Escenario 1
#-----
simula_bb<-read.table("Escenario1.txt",header=TRUE)
x<-seq(1,70,2);x
y<-seq(2,70,2);y
Esc1E<- simula_bb[x,3]; Esc1E
Esc1C<- simula_bb[y,3]; Esc1C
sigma<- simula_bb[,4]; sigma
sim<- data.frame(Esc1C,Esc1E,sigma); sim

disp1 <- ggplot(sim, aes(Esc1C,Esc1E,color=sigma))+
  geom_point()+
  labs(color =c(expression(Sigma)))+
  theme(axis.title.x = element_blank(),
        axis.title.y = element_blank())
print(disp1)
#-----

```

```

#Escenario 2
#-----

simula_bb2<-read.table("Escenario2.txt",header=TRUE)
x<-seq(1,70,2)
y<-seq(2,70,2)
Esc2E<- simula_bb2[x,3]
Esc2C<- simula_bb2[y,3]
sigma<- simula_bb2[ ,4]
sim2<- data.frame(Esc2E,Esc2C,sigma); sim2

disp2 <- ggplot(sim2, aes(Esc2C,Esc2E,color=sigma))+
  geom_point()+
  labs(color =c(expression(Sigma)))+
  theme(axis.title.x = element_blank(),axis.title.y = element_blank())
print(disp2)
#-----

##Análisis descriptivo (Capítulo 4)
#-----
#-----

#Base de datos (https://sinca.mma.gob.cl)
#-----

#Tabla 4.1
#Figura 4.1
#Figura 4.2
APLICACION_TESIS<- read_excel("F:./DATOS.xlsx",sheet = "Mayo")
MP25<- APLICACION_TESIS[3]; (MP25)
MP10<- APLICACION_TESIS[4]; (MP10)
#-----

#MP2.5
#-----

MP25<- as.numeric(as.character(unlist(MP25[[1]])))
Hora<- as.numeric(as.character(unlist(APLICACION_TESIS[,2] [[1]])))

```

```

Dia<- as.numeric(as.character(unlist(APLICACION_TESIS[,1][[1]])))
layout(matrix(c(1, 1, 1, 1, 2, 3, 2, 3), nrow=4, byrow=TRUE))
plot(MP25,type = "p", col = "grey90",main = "MP2.5")
boxplot(MP25, col="grey90",bty="n")
legend("bottomleft", legend = c("MP2.5") ,bty = "n", col = "grey90",
      pch=20,pt.cex=3,cex=1,horiz=FALSE,inset=c(0.03,0.1))
hist(MP25, col="grey90", main="",lty=9)
summary(MP25)
var(MP25)
sd(MP25)
sd(MP25)/mean(MP25)
kurtosis(MP25)
skewness(MP25)
#-----
#MP10
#-----
MP10<- as.numeric(as.character(unlist(MP10[[1]])))
layout(matrix(c(1, 1, 1, 1, 2, 3, 2, 3), nrow=4, byrow=TRUE))
plot(MP10,type = "p", col = "grey90",main = "MP10")
boxplot(MP25, col="grey90",bty="n")
legend("bottomleft", legend = c("MP10") ,bty = "n", col = "grey90",
      pch=20,pt.cex=3,cex = 1,horiz = FALSE,inset =c(0.03, 0.1))
hist(MP25, col="grey90", main="",lty=9)
summary(MP25)
summary(MP10)
var(MP10)
sd(MP10)
sd(MP10)/mean(MP10)
kurtosis(MP10)
skewness(MP10)
#-----
cor(MP25,MP10)

```

```

ggpairs(APLICACION_TESIS[,c(3,4)])
ggplot(data = APLICACION_TESIS)+geom_point(mapping = aes(x=MP25,y=MP10))
ggpairs(APLICACION_TESIS[,c(3,4)], diag = list(continuous = "bar"))
#-----
#-----
#Cartas de control Univariadas
#-----
#-----
#Para MP2.5
#-----
#Cargamos nuestra base de datos
MP25<- read_excel("F:/DATOS.xlsx",sheet = "MayoMP25");
#Se hallan las desviaciones, S
desviaciones<-c(apply(MP25, 2, sd)) ; desviaciones
#Se halla S Barra
SBarra<-round(mean(desviaciones),2); SBarra
#Eje de las X's
dia<-1:ncol(MP25)
#Se establecen las constantes a utilizar, en este caso para n=24
B4<-1.445
B3<-0.555
#Se crean los límites de control
LCS<- round(B4*SBarra,2); LCS
LC <- round(SBarra,2) ;LC
LCI<- round(B3*SBarra,2); LCI
#Se realiza el gráfico de control
par(mfrow=c(1,1))
plot(dia,desviaciones,type="l",xaxt="n",pch=19,
      xlim=c(0,(length(desviaciones)+6)),
      ylim=c(min(LCI,min(desviaciones))-1,
              max(LCS,max(desviaciones))+1))
#Se realiza el gráfico de control

```

```

#Textos agregados
axis(1,1:(length(desviaciones)),cex.axis=0.6,las=2)
abline(h=LC,lty=3,col="green")
text((max(dia)+1),SBarra,paste("S=",SBarra),pos=1,font=2,cex=0.6)
abline(h=LCS,lty=3,col="red")
text((max(dia)+1),LCS,paste("LCS=",LCS),pos=1,font=2,cex=0.6)
abline(h=LCI,lty=3,col="red")
text((max(dia)+1),LCI,paste("LCI=",LCI),pos=1,font=2,cex=0.6)
#Condición para cuando existan valores fuera de control
for(i in 1:length(desviaciones))
  if(desviaciones[i]>LCS)temp<-4
  else if(desviaciones[i]<LCI)temp<-4
  else temp<-19
  points(dia[i],desviaciones[i],pch=temp)

#Texto adicional
mtext("Carta de control S",side=3,font=2)
#-----
#Para MP10
#-----
#Cargamos nuestra base de datos
MP10<- read_excel("F:/./DATOS.xlsx",sheet = "MayoMP10")
#Se hallan las desviaciones, S
desviaciones<-c(apply(MP10, 2, sd)) ; desviaciones
#Se halla S Barra
SBarra1<-mean(desviaciones); SBarra1
#Eje de las X's
dia<- 1:ncol(MP10)
#Se establecen las constantes a utilizar, en este caso para n=24
B4<-1.445
B3<-0.555
#Se crean los límites de control

```

```

LCS<- round(B4*SBarra1,2); LCS
LC<- round(SBarra1,2) ;LC
LCI<- round(B3*SBarra1,2); LCI
#Se realiza el gráfico de control
plot(dia,desviaciones,type="l",xaxt="n",pch=19,
      xlim=c(0,(length(desviaciones)+2)),
      ylim=c(min(LCI,min(desviaciones))-0.1,
              max(LCS,max(desviaciones))+0.1))
#Textos agregados
axis(1,1:(length(desviaciones)+2),cex.axis=0.6,las=2)
abline(h=LC,lty=3,col="green")
text((max(dia)+1),SBarra1,paste("S=",LC),pos=1,font=2,cex=0.6)
abline(h=LCS,lty=3,col="red")
text((max(dia)+1),LCS,paste("LCS=",LCS),pos=1,font=2,cex=0.6)
abline(h=LCI,lty=3,col="red")
text((max(dia)+1),LCI,paste("LCI=",LCI),pos=1,font=2,cex=0.6)
#Condición para cuando existan valores fuera de control
for(i in 1:length(desviaciones))
  if(desviaciones[i]>LCS)temp<-4
  else if(desviaciones[i]<LCI)temp<-4
  else temp<-19
  points(dia[i],desviaciones[i],pch=temp)

#Texto adicional
mtext("Carta de control S",side=3,font=2)
#-----
#Carga de control Mutivarida, $T2$ de Hotelling
#-----
x<- read_excel("F:./DATOS.xlsx",sheet = "MayoDiaria")
x<- as.data.frame(x)
str(x)
data.mqcd <- mqcd(x)

```

```

res.mqcs <- mqcs.t2(data.mqcd)
summary(res.mqcs)
plot(res.mqcs, title ="Carta de control Hotelling")
#-----
  x<- as.matrix(x)
carta.T2<-function(x,alpha)
  n<-nrow(x)
  p<-ncol(x)
  LC<-round((((n-1)^2)/n)*qbeta((1-alpha),(p/2),(n-p-1)/2),2)
  medias.x<-apply(x,2,mean)
  medias<-t(medias.x)
  uno<-c(rep(1,nrow(x)))
  matriz.uno<-matrix(t(uno))
  matriz.medias<-matriz.uno%*%medias
  datos.centrados<-x-matriz.medias
  matriz.covarianza<-var(x)
  inversa<-solve(matriz.covarianza)
  aux<-datos.centrados%*%inversa%*%t(datos.centrados)
  T2<-diag(aux)
  tiempo<-1:length(T2)
  plot(tiempo,T2,type="l",xlim=c(0,(length(T2)+1)),
       ylim=c(0,max(LC,max(T2))+2),xaxt="n")
  axis(1,1:(length(T2)+1),cex.axis=0.8,las=2)
  abline(h=LC,lty=3, col="red")
  for(i in 1:length(T2))temp<-ifelse((T2[i]>LC),4,19)
  points(tiempo[i],T2[i],pch=temp)

  text((max(tiempo)-1),LC,paste("LC=",LC),pos=3,font=2,cex=0.8)
  mtext(paste("Carta T2","alpha=",alpha),side=3,font=2)

carta.T2(x,0.005)
#-----

```

```

#-----
#Carta de control |S|
#-----
#-----
ejCMV<-read_excel("F:./DATOS.xlsx",sheet = "Mayo");ejCMV
# vector de medias
Xbb<-apply(ejCMV[3:4],2,mean);Xbb
# numero de grupos
m<-length( table(ejCMV$Fecha));m
p<-length(Xbb);p
n<-nrow(ejCMV[ejCMV$Fecha==1,3:4]);n
#Para calcular S barra
for(i in 1:m)
  if(i==1)
    varcov<-cov(ejCMV[ejCMV$Fecha==i,3:4])
    Xb<-apply(ejCMV[ejCMV$Fecha==i,3:4],2,mean)
    detS<-det( cov(ejCMV[ejCMV$Fecha==i,3:4]) )
  else
    varcov<-varcov+cov(ejCMV[ejCMV$Fecha==i,3:4] )
    Xb<-rbind(Xb, apply(ejCMV[ejCMV$Fecha==i,3:4],2,mean))
    detS<-c(detS,det( cov(ejCMV[ejCMV$Fecha==i,3:4]) ))

# fin del for
# matriz S barra
Sb<-varcov/m;Sb
# matriz de vectores de media
Xb
#determinantes para ubicar
detS
# las constantes b1 y b2
b1<-(1/(n-1)^p)*prod(n-(1:p))
b2<-(1/(n-1)^(2*p))*prod(n-(1:p) )*(prod(n-(1:p)+2)-prod(n-(1:p)))

```

```

# estimador de determinante de sigma
detSo<-det(Sb)/b1
LCS<-det(Sb)*( 1 + 3*(sqrt(b2)/b1) );LCS
LC<-det(Sb);LC
LCI<-det(Sb)*( 1 - 3*(sqrt(b2)/b1) );LCI
if(LCI<0)LCI<-0

plot(dia,detS,type="l",xaxt="n",pch=19,ylab="|S|",
      xlim=c(0,(length(detS)+2)),
      ylim=c(min(LCI,min(detS))-0.1,
              max(LCS,max(detS))+0.1))
#Textos agregados
axis(1,1:(length(detS)),cex.axis=0.6,las=2)
abline(h=LC,lty=3,col="green")
text((max(dia)+1),LC,paste("LC=",LC),pos=3,font=2,cex=0.4)
abline(h=LCS,lty=3,col="red")
text((max(dia)+1),LCS,paste("LCS=",LCS),pos=3,font=2,cex=0.4)
abline(h=LCI,lty=3,col="red")
text((max(dia)+1),LCI,paste("LCI=",LCI),pos=3,font=2,cex=0.4)
#Condición para cuando existan valores fuera de control
for(i in 1:length(detS))
  if(detS[i]>LCS)temp<-4
  else if(detS[i]<LCI)temp<-4
  else temp<-19
  points(dia[i],detS[i],pch=temp)
#-----
#Aplicación bajo simulación
#-----
#options(max.print=5000)
#options(max.print = .Machine$integer.max)
#Se usa la función antes creada Resultado(n,m,K) adicionando la lectura
a<- data.frame(Resultado1, RL,LCI1,LC1,LCS1,Resultado2,RL2,LCI2,LC2,LCS2)

```

```

write.table(a, "ResultadosAplicacionSimulacion10005.txt", sep="^",quote =F,
row.names = F)
x<- read_excel("F:./DATOS.xlsx",sheet = "MayoDiaria")
x<- as.data.frame(x)
sink("ResultadosAplicacionSimulacion10005(5,30,1000)2.txt")
Resultado(5,30,1000)
#Resultado(1000,30,1000)
#Resultado(5,1000,1000)
#Resultado(1000,1000,1000)
sink()
#-----
#Gráfico Bootstraps (Figura 4.7)
#-----
setwd("F:./Tesis Salidas .txt")
x<- x[c(1:100),1]
##20
resultados<- rep(0,1000)
for (i in 1:1000)resultados[i]<- mean(x[sample(1:length(x),20,replace=T)])
###
datos$muestra=factor(datos$muestra,levels =c("m=20","m=40","m=60","m=80",
"m=100"))
levels(datos$muestra)
gr4 <- ggplot(datos, aes(x=R)) + geom_histogram()
gr4 + facet_grid(.~ muestra)
###End
#-----

```