



**ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL**

**Facultad de Ciencias Naturales y Matemáticas**

**TRABAJO FINAL DE LA MATERIA INTEGRADORA**

**“Diseño de gráficos de control  
multivariantes basado en la combinación  
lineal de variables normales”**

Previo a la obtención del Título de:

**INGENIERO EN ESTADÍSTICA INFORMÁTICA**

Presentado por:

Andrés Eduardo Plaza Rodas

GUAYAQUIL – ECUADOR

Año: 2017

## **AGRADECIMIENTO**

A Dios por la fortaleza y sabiduría necesaria, para llevar a cabo este proyecto de investigación. A mis padres, por su apoyo incondicional, enseñanza y guía a lo largo de mi vida universitaria.

A mi tutora de tesis la Ph.D. Sandra García, por su paciencia, dedicación y conocimientos necesarios para la realización de este proyecto. A mis compañeros y amigos con los que compartí varias experiencias en el ámbito estudiantil y académico.

Finalmente, a todos los profesores de mi carrera universitaria por su ardua labor y dedicación en su trabajo.

## **DEDICATORIA**

A Dios, a mi familia, especialmente a mis abuelos (†) y a mi compañero, amigo y profesor el Mg. Elkin Angulo (†).

## **DECLARACIÓN EXPRESA**

"La responsabilidad del contenido desarrollado en la presente propuesta de la materia integradora corresponde exclusivamente a Andrés Eduardo Plaza Rodas y el patrimonio intelectual del mismo a la ESCUELA SUPERIOR POLITÉCNICA DEL LITORAL"

---

Andrés Eduardo Plaza Rodas

## RESUMEN

Utilizar un esquema de control multivariante a través de un estadístico que involucre a todas las características de calidad de un producto, es una alternativa al monitorear múltiples variables especialmente cuando están correlacionadas. Cuando las variables son normales la estrategia multivariante más común es el uso del gráfico de control  $T^2$  de Hotelling; sin embargo, este gráfico muestra insensibilidad cuando existen cambios pequeños en el vector de medias a medida que el número de variables  $p$ , se incrementa. Para minimizar dichos inconvenientes, en este proyecto se analizó el control simultáneo de un conjunto de variables normales correlacionadas a través de dos gráficos de control propuestos: el primero denominado LCN, que utiliza un estadístico fundamentado en la combinación lineal de variables normales que considera sus correlaciones; y el segundo, denominado LCPC, sustentado en la combinación lineal de al menos las dos primeras componentes principales, que expliquen como mínimo el 80% de la variabilidad total. Uno de los objetivos de esta investigación es analizar el desempeño de estos gráficos. Para ello, se desarrolló un software que encuentra los mejores parámetros a través de una técnica de optimización heurística denominada Algoritmos Genéticos (GA); de manera que, se minimice el ARL fuera de control para un cambio propuesto en el vector de medias. De esta manera, se detectará el cambio en el proceso lo más rápido posible, restringido a un valor de ARL bajo control deseado. Los gráficos de control propuestos son de tipo Shewhart y presentan mejores rendimientos que el gráfico de control  $T^2$  de Hotelling.

**Palabras claves:** Gráficos de control, Variables normales, Algoritmos Genéticos.

## **ABSTRACT**

*Using a multivariate control scheme through a statistic that involves all the quality characteristics of a product, is an alternative to monitoring multiple variables especially when they are correlated. When we work with normal variables the most common multivariate strategy is the use of Hotelling  $T^2$  control chart; however, this chart shows insensitivity when there are small changes in the means vector as the number of variables  $p$  increases. To minimize such drawbacks, this project analyzed the simultaneous control of a set of normal variables correlated through two proposed control charts: the first one called LCN, which uses a statistic based on the linear combination of normal variables that considers their correlations; and the second, called LCPC, based on linear combination of at least two principal components, explaining at least 80% of the total variability. As one of the objectives of this research is to analyze the performance of these charts, it developed a friendly program that finds the best parameters through a heuristic called Genetic Algorithms (GA), so that the out-of-control ARL (Average Run Length) is minimized for a proposed shift in the mean vector to be detected as quickly as possible restricted to a desired in-control ARL value. The proposed control charts are Shewhart type and show better performance than the Hotelling  $T^2$  control chart.*

**Key words:** Control charts, Normal variables, Genetic algorithms.

# ÍNDICE GENERAL

RESUMEN .....	I
ABSTRACT .....	II
ÍNDICE GENERAL.....	VI
ACRÓNIMOS O SIGLAS .....	IX
GLOSARIO DE TÉRMINOS Y SIMBOLOGÍAS .....	X
ÍNDICE DE FIGURAS .....	XI
ÍNDICE DE TABLAS .....	XII
CAPÍTULO 1 .....	1
1. INTRODUCCIÓN .....	1
1.1. Planteamiento del problema .....	1
1.2. Antecedentes .....	3
1.3. Motivación.....	4
1.4. Objetivos .....	5
1.4.1.Objetivo General.....	5
1.4.2.Objetivos Específicos .....	5
1.5. Alcance de estudio.....	6
1.6. Marco Teórico .....	7
1.6.1.Contraste de Hipótesis .....	7
1.6.2.Gráficos de control.....	8
1.6.3.Gráfico de control $T^2$ de Hotelling .....	17
1.6.4.Algoritmos Genéticos (GA) .....	19

1.7. Revisión Bibliográfica.....	24
1.7.1. Selección de variables fuera de control con el procedimiento de calidad multivariante $T^2$ .....	24
1.7.2. Control estadístico de procesos multivariados.....	25
1.7.3. Monitoreo y control en un proceso normal multivariado .....	26
1.7.4. Combinación lineal óptima de variables Poisson para el control estadístico multivariante de procesos. ....	26
CAPÍTULO 2 .....	28
2. METODOLOGÍA DEL DISEÑO.....	28
2.1. Gráfico de control basado en la combinación lineal de variables normales correlacionadas (LCN control chart). ....	28
2.2. Gráfico de control basado en la combinación lineal de componentes principales (LCPC control chart).....	30
2.3. Problema de optimización para el gráfico de control LCN y LCPC .....	32
CAPÍTULO 3 .....	34
3. RESULTADOS.....	34
3.1. Software y ejemplo de aplicación.....	34
3.2. Comparación de rendimientos .....	41
3.3. Análisis de sensibilidad.....	43
CAPÍTULO 4 .....	51



4. DISCUSIÓN Y CONCLUSIONES .....	51
4.1. Conclusiones .....	51
4.2. Recomendaciones .....	52
BIBLIOGRAFÍA .....	53
APÉNDICES .....	57
APÉNDICE A.....	57
APÉNDICE B.....	58
APÉNDICE C.....	59
APÉNDICE D.....	60
APÉNDICE E.....	61
APÉNDICE F .....	62
APÉNDICE G .....	63
APÉNDICE H.....	64
APÉNDICE I .....	65

## ACRÓNIMOS O SIGLAS

SPC	Control Estadístico de Procesos
LCN	Gráfico multivariante de la combinación lineal de variables normales
LCPC	Gráfico multivariante de la combinación lineal de componentes principales
LCS	Límite de control superior
LCI	Límite de control inferior
LC	Línea central
ARL	Longitud promedio de rachas
ARL <sub>0</sub>	ARL bajo control deseado
ARL <sub>1</sub>	ARL fuera de control
ARL(d=0)	ARL bajo control obtenido por el software
ARL(d=d*)	ARL fuera de control obtenido por el software para un cambio d*
d*	Vector promedio de cambios en sigma unidades
GA	Algoritmos Genéticos
n	Tamaño de la muestra
m	Número de muestras por lote
p	Número de variables a controlar

## GLOSARIO DE TÉRMINOS Y SIMBOLOGÍAS

Algoritmo	Conjunto de instrucciones ordenadas y programadas en un software que permiten hallar la solución a un determinado problema.
Control estadístico de procesos	Utilización de gráficos de control u otras técnicas estadísticas que permitan controlar la variabilidad de un proceso.
Correlación	Establece la relación en magnitud y dirección lineal entre dos variables.
Distribución de probabilidad	Función matemática que asigna la probabilidad de suceso de una variable aleatoria para un valor determinado.
Estadístico	Función medible con distribución de probabilidad, que permite inferir o estimar características de un proceso a través de una
Muestra	Representación del proceso en un subgrupo o subconjunto de mediciones, escogidas por algún método de muestreo.
Multivariante / Multivariado	Análisis de dos o más variables en algún método o técnica estadística.
Variables de calidad	Características de interés a controlar en un proceso, los cuales pueden ser medibles y observables.

## ÍNDICE DE FIGURAS

Figura 1.1. Estructura de un gráfico de control .....	11
Figura 1.2. Fases en la ejecución de un algoritmo genético .....	21
Figura 3.1. Salida del software para el gráfico de control LCN optimizado para el ejemplo de aplicación propuesto.....	36
Figura 3.2. Gráfico de control LCN para un proceso de horneado fuera de control, con datos simulados. ....	37
Figura 3.3. Salida del software para el gráfico de control LCPC optimizado para el ejemplo de aplicación propuesto.....	39
Figura 3.4. Gráfico de control LCPC para un proceso de horneado fuera de control, con datos simulados. ....	40

## ÍNDICE DE TABLAS

Tabla 1. Tipos de errores en un contraste de Hipótesis.....	8
Tabla 2. Parámetros más comunes a controlar y sus estimadores.....	13
Tabla 3. Estructura de los gráficos de control univariados más usuales.....	14
Tabla 4. Conceptos básicos utilizados en algoritmos genéticos (GA).....	20
Tabla 5. Escenarios analizados en la comparación de rendimientos de los gráficos de control diseñados.....	41
Tabla 6. Vectores de cambios considerados para el análisis de sensibilidad del gráfico de control LCN. ....	44
Tabla 7. ARLs fuera de control para los gráficos LCN optimizados para los vectores de cambios: $a = (0.25, 0.25)$ , y $b = (2.00, 2.00)$ ; con una correlación positiva de 0.85 y $ARL_0 = 465$ . Los valores entre paréntesis son las mediciones obtenidas para el cambio (b). ....	45
Tabla 8. ARLs fuera de control para los gráficos LCN optimizados para los vectores de cambios: $c = (-0.25, -0.25)$ , y $d = (-2.00, -2.00)$ ; con una correlación positiva de 0.85, $ARL_0 = 465$ . Los valores entre paréntesis son las mediciones obtenidas para el cambio (d). ....	46
Tabla 9. ARLs fuera de control para los gráficos LCN optimizados para los vectores de cambios: $e = (-1.20, -2.00)$ , y $f = (-0.60, -1.20)$ ; con una correlación positiva de 0.35, $ARL_0 = 750$ . Los valores entre paréntesis son las mediciones obtenidas para el cambio (f). ....	47
Tabla 10. ARLs fuera de control para los gráficos LCN optimizados para los vectores de cambios: $g = (-0.25, 0.50)$ , y $h = (-0.75, 1.25)$ ; con una correlación negativa de -0.45, $ARL_0 = 900$ . Los valores entre paréntesis son las mediciones obtenidas para el cambio (h).....	48

Tabla 11. ARLs fuera de control para los gráficos LCN optimizados para los vectores de cambios:  $i = (-0.90, 0.90)$ ,  $y j = (-1.50, 1.50)$ ; con una correlación negativa de  $-0.85$ ,  $ARL_0 = 500$ . Los valores entre paréntesis son las mediciones obtenidas para el cambio ( $j$ ). .....49

# CAPÍTULO 1

## 1. INTRODUCCIÓN

La presente investigación analiza el control simultáneo de un conjunto de variables normales correlacionadas a través de dos gráficos de control propuestos, denominados LCN y LCPC.

Primero, se expondrá el planteamiento del problema, sus antecedentes, la motivación que se tuvo para llevar a cabo la investigación, los objetivos que se desean alcanzar y el alcance de estudio. Además, se presentará el marco teórico necesario para la elaboración de este proyecto y el resumen de algunas investigaciones acerca del control estadístico multivariante.

El Capítulo 2 expone la metodología aplicada para esta investigación, explicando a fondo cada uno de los gráficos de control multivariantes a comparar con el  $T^2$  de Hotelling y a su vez se expone el problema de optimización planteado.

Un ejemplo de aplicación del software desarrollado en esta investigación, la comparación de rendimientos entre los distintos gráficos de control analizados y el análisis de sensibilidad de la carta de control LCN se detallan en el Capítulo 3.

Finalmente, se presentan las conclusiones y recomendaciones de este proyecto de investigación.

### 1.1. Planteamiento del problema

Los gráficos de control son herramientas muy utilizadas en el monitoreo estadístico de las variables de calidad, y al ser estas, parte de un proceso

o servicio, tienden a estar correlacionadas en la mayoría de los casos, por lo que, resulta pertinente considerar esta estructura de correlación en el control estadístico efectuado sobre estas variables. Para controlar varias variables de calidad a la vez, por simplicidad se puede utilizar un esquema múltiple, es decir un gráfico de control para cada variable a monitorear, sin embargo, este método tiene asociado dos inconvenientes, el primero es que “la probabilidad del error tipo I de este esquema sería mayor a la considerada en el diseño de cada gráfico univariante” (Cook & Chiu, 1998) y el segundo es que no suele considerar las correlaciones de las variables, resultando inadecuado su uso en el caso de que las variables estén correlacionadas.

Ante estos problemas se tiene otra alternativa de control para varias variables, el de usar un esquema multivariante, un solo gráfico de control con un estadístico que involucre a todas las variables de calidad; el cual, “tiene mayor sensibilidad cuando existen cambios en el vector de medias, ya que, considera la estructura de correlación de las variables originales, siendo esta, la idea esencial en el uso de este tipo de cartas de control” (Aparisi, de Luna, & Epprecht, 2010). Lowry y Montgomery (1995), resaltan que “la elección de la mejor alternativa dependerá del número de variables, la estructura de correlación y la dirección del cambio en el vector de medias”.

Por lo tanto, para minimizar los problemas mencionados, esta investigación está enfocada en diseñar un gráfico de control que monitoree un conjunto de variables normales junto a sus correlaciones y además que mejore el rendimiento de la carta de control propuesta por Harold Hotelling en 1947.



## 1.2. Antecedentes

En el área de control estadístico multivariante de calidad hay varios trabajos desarrollados. Debido a la utilidad del Teorema de Límite Central y a las propiedades de las variables aleatorias normales, las investigaciones relacionadas a esquemas multivariantes o múltiples para variables normales son muchas, siendo las más destacadas, la del gráfico de control  $T^2$  de Hotelling (Hotelling, 1947); la del gráfico denominado Multivariate Exponentially-Weighted Moving Average (Lowry et al., 1992) y la de la carta de control denominada Multivariate Cumulative Sum (Pignatiello & Runger, 1990; Crosier, 1998). Otros autores también desarrollaron gráficos multivariantes o metodologías de control de calidad basadas en distribuciones normales como Murphy (1987), quien propuso una prueba para seleccionar variables fuera de control y una interpretación a los valores del estadístico  $T^2$  bajo una distribución normal multivariante. Shabbak y Midi (Shabbak & Midi, 2012) propusieron una mejora del gráfico de control  $T^2$  considerando un límite de control superior basado en la mediana y su desviación absoluta. Tran y otros autores (Tran, Castagliola, & Celano, 2016) desarrollaron una metodología para monitorear la proporción de dos variables normales usando reglas complementarias que, para evaluar el rendimiento de este esquema, los autores utilizaron Cadenas de Markov y una aproximación normal eficiente de la distribución de dicha proporción.

En cuanto a esquemas múltiples, se tiene el trabajo desarrollado por Hayter y Tsui (Hayter & KNOW-LEUNG, 1994), quienes propusieron un procedimiento de control para las medias de las variables basado en intervalos de confianza simultáneos, utilizando la distribución normal multivariante se determinan los puntos críticos de estos intervalos. Por otra parte, se han desarrollado esquemas de control que consideran estadísticos como combinaciones lineales de variables aleatorias, tal es el caso de los gráficos de control basados en componentes principales.

MacGregor y Kourti (MacGregor & Kourti, 1995) diseñaron gráficos de control multivariados para monitorear procesos por lotes utilizando el análisis de componentes principales de múltiples vías. Otro aporte lo hacen Zamarrón et al. (Zamarrón, Prado, & Luis, 2012), quienes propusieron una metodología integrada por la gráfica  $T^2$  Hotelling que detecta señales fuera de control y las componentes principales para identificar qué variable es la causante del estado fuera de control. Además, otros trabajos relacionados al control estadístico multivariante mediante la combinación lineal de las variables de calidad, son los propuestos por Epprecht et al. (Epprecht, García-Bustos, & Aparisi, 2013), García-Bustos et al. (García-Bustos, Aparisi, & Epprecht, Optimal EWMA of Linear Combination of Poisson Variables for Multivariate Statistical Process Control, 2014) y García-Bustos et al. (García-Bustos, Mite, & Vera, Control Charts with Variable Dimension for Linear Combination of Poisson Variables., 2015), quienes construyeron gráficos de control óptimos para variables Poisson correlacionadas basados en combinaciones lineales de estas variables discretas.

### **1.3. Motivación**

Como se puede observar en los antecedentes, el control de calidad multivariante para variables continuas está muy desarrollada; donde por lo general, se trata de diseñar gráficos de control que tengan una baja tasa de falsas alarmas y detecten rápidamente los cambios que suelen darse en el proceso. La medida de rendimiento mayormente utilizada en gráficos de control es aquella denominada ARL (Average Run Length), el número promedio de muestras que se toman hasta que se presenta una señal. Para lo cual, el diseño óptimo de un gráfico de control es aquel que posee un valor grande de ARL cuando el proceso está bajo control y un ARL mínimo posible, cuando el proceso está fuera de control.

Las industrias por facilidad, suelen utilizar esquemas múltiples para el control de calidad de la mayoría de sus procesos, esto puede conllevar a varios problemas si las variables a controlar están correlacionadas, donde para estos casos, el esquema de control multivariante más utilizado es el  $T^2$  de Hotelling que tiene algunos inconvenientes en detectar pequeños cambios en el proceso, rápidamente. Esto motivó a diseñar dos gráficos de control optimizados: el primero denominado LCN, que utiliza un estadístico fundamentado en la combinación lineal de variables normales que considera sus correlaciones; y el segundo, denominado LCPC, sustentado en la combinación lineal de al menos las dos primeras componentes principales, que expliquen como mínimo el 80% de la variabilidad total.

#### **1.4. Objetivos**

##### **1.4.1. Objetivo General**

Diseñar gráficos de control óptimos sustentados en la combinación lineal de variables normales, que superen el rendimiento de la carta de control  $T^2$  de Hotelling.

##### **1.4.2. Objetivos Específicos**

- Desarrollar un estadístico basado en la combinación lineal de variables normales.
- Desarrollar un estadístico fundamentado en la combinación lineal de al menos las dos primeras componentes principales, que expliquen como mínimo el 80% de la variabilidad total.
- Realizar la programación de algoritmos genéticos en Visual Basic Community© que permitan obtener los coeficientes de los estadísticos propuestos, de manera que sea posible alcanzar un

ARL bajo control deseado y a su vez, minimizar el ARL fuera de control para un cambio dado en los parámetros del proceso.

- Validar mediante un ejemplo de aplicación con datos simulados, la mejora en el rendimiento de los estadísticos propuestos, en comparación con el  $T^2$  de Hotelling.
- Realizar la comparación de rendimientos, entre los gráficos de control: LCN, LCPC y  $T^2$  de Hotelling, considerando valores diferentes del  $ARL_0$ , diversos cambios en las medias y distintas correlaciones del conjunto de variables a controlar.
- Realizar un análisis de sensibilidad en el gráfico de control basado en la combinación lineal de variables normales y así, conocer su rendimiento ante cualquier cambio inesperado en el proceso.

### **1.5. Alcance de estudio**

Los gráficos de control analizados en este proyecto son de tipo Shewhart y sólo se basan en variables normales correlacionadas bajo la relación lineal. A fin de cuantificar esta relación se utilizó el coeficiente de correlación definido por Pearson.

Como se mencionó anteriormente, dichos gráficos de control están basados en la combinación lineal de variables normales, en las que sus parámetros han sido optimizados utilizando Algoritmos Genéticos que, a su vez, minimiza el ARL fuera de control para un cambio en los parámetros del proceso, restringido a un ARL bajo control deseado que es fijado por el usuario final. El vector de cambios a analizar, está medido en unidades de desviación estándar con respecto a las medias de las variables normales involucradas en el proceso.

## 1.6. Marco Teórico

### 1.6.1. Contraste de Hipótesis

Una hipótesis, estadísticamente hablando (Wackerly, Mendenhall, & Scheaffer, 2010), es un supuesto que se plantea respecto a los parámetros de una distribución de probabilidad de una población, por lo tanto, el contraste será el método o procedimiento para analizar si ésta debe ser rechazada o no. Se distinguen dos tipos:

- La  $H_0$  o hipótesis nula, establecido al principio de toda investigación con el fin de poder rechazarla o no.
- La  $H_a$  o hipótesis alterna, complementaria a la anterior.

Se parte del supuesto que  $H_0$  es verdadera, para contrastar una hipótesis nula contra la alterna.

#### 1.6.1.1. Posibles errores en un contraste de Hipótesis

Al establecer un contraste de hipótesis, se define un criterio que permite decidir si dicho supuesto se rechaza o no y, por lo tanto, no se establece su verdad absoluta ya que, las muestras analizadas difieren significativamente de los resultados esperados.

Por consiguiente, según sea la situación real y la decisión que se tome se cometen dos tipos de errores: el tipo I con probabilidad  $\alpha$ , si se rechaza la hipótesis nula en caso de ser cierta, y el tipo II con probabilidad  $\beta$ , si se acepta  $H_0$  cuando debe ser rechazada. La siguiente tabla resume lo expuesto:

**Tabla 1. Tipos de errores en un contraste de Hipótesis**

	<b>Decisiones</b>	
<b>Eventos</b>	No rechazar $H_0$	Rechazar $H_0$
$H_0$ es verdadera	Correcto	Error tipo I $P(\text{Error tipo I}) = \alpha$
$H_0$ es falsa	Error tipo II $P(\text{Error tipo II}) = \beta$	Correcto

*Fuente: Elaboración propia*

## **Gráficos de control**

### **1.6.2.1. Generalidades**

El Dr. Walter Shewhart mientras trabajaba en la empresa Bell Telephone, preparó las bases del Control Estadístico de Procesos, estableciendo así la idea del diseño de gráficos de control en el año 1924. Propuso que para asegurar la calidad en un proceso era necesario llevarlo al estado bajo control y a su vez que la variabilidad se reduzca, existiendo dos fuentes: la originada por causas aleatorias o especiales. Por lo tanto, el gráfico de control propuesto por Shewhart, constituye una herramienta para identificar las causas aleatorias y especiales de variabilidad que se encarga de monitorear una característica de calidad cuantitativa o cualitativa a través de límites de control y un estadístico muestral.

Si su valor cae dentro de los límites, se considera que el proceso se encuentra bajo control caso contrario, podría encontrarse fuera de control, donde debería estudiarse si la situación se debe a una causa especial o aleatoria. Si la causa es especial, se deben realizar procedimientos que lleven al proceso al estado bajo control (Montgomery, 2007).

Shewhart mencionó que el control estadístico de procesos consta de dos fases:

- La fase de desarrollo, que determina los parámetros del proceso, los límites de control, se establece un estimador y se eliminan causas asignables o especiales.
- La fase de madurez, donde se identifica el número de muestras necesarias para detectar pequeños cambios en el proceso y el número promedio de muestras, antes de obtener falsas alarmas.

Los primeros gráficos de control propuestos por Shewhart, monitoreaban variables cuantitativas utilizando como estadísticos muestrales el promedio  $\bar{X}$  y la variación, ya sea por rango R o por desviación estándar S. Luego, se desarrollaron gráficos de control para controlar variables cualitativas, como el gráfico p y np, que sirven para monitorizar la proporción y el número de unidades no conformes por lotes; el gráfico c y u que monitorizan el número y proporción de defectos por lote, que detectan eficientemente grandes cambios en el proceso.

Los gráficos de control univariados CUSUM y EWMA, diseñados en 1954 y 1956 por E. S. Page y S. W. Roberts, se encargan de detectar pequeños cambios en el proceso y consideran la información muestral histórica, no sólo el último valor como lo hacen los gráficos Shewhart.

El primero se basa en las sumas acumuladas de las desviaciones de los valores muestrales respecto a un valor objetivo y el segundo considera los promedios móviles exponencialmente ponderados.

Además, con el fin de controlar varias variables de calidad con su estructura de correlación, se desarrollaron estadísticos multivariantes como el propuesto en 1947 por Harold Hotelling, quien presenta un procedimiento para controlar el vector de medias de una población normal multivariada, mediante el estadístico  $T^2$  que mide la distancia entre el vector de medias esperado y el vector de medias observado tomando en cuenta su matriz de varianzas y covarianzas. Se detallará mucho más acerca de este gráfico en el apartado 1.6.3.

Otros estadísticos multivariantes que consideran la estructura de correlación de un conjunto de variables, son los propuestos por Crossier (Crossier, 1988) y Lowry et al. (Lowry, Woodall, Champ, & Rigdon, 1992), quienes presentan una versión multivariada del gráfico de control de sumas acumuladas (MCUSUM) y de los promedios móviles exponencialmente ponderados (MEWMA).

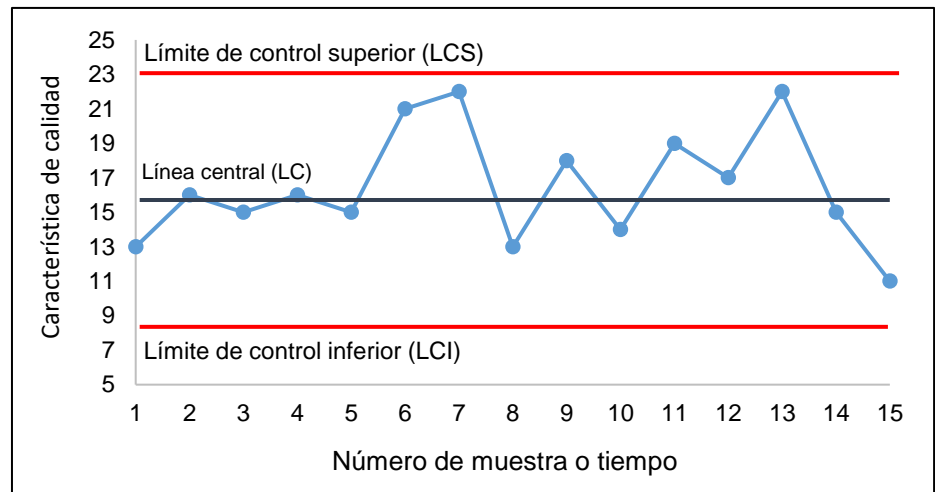
#### **1.6.2.2. Definición**

Una gráfica o carta de control está definido como “la representación de funciones estadísticas de las mediciones de una o varias características de calidad en muestras tomadas del proceso, en función del tiempo o del número de muestra” (Montgomery, 2007). Forma parte de una de las técnicas primarias del Control Estadístico de Procesos (SPC) y son muy útiles para garantizar la calidad del producto final, a través de la disminución de la variabilidad (Lowry & Montgomery, A review of multivariate control charts., 1995).

Generalmente este esquema incluye límites de control superior (LCS) e inferior (LCI), además de una línea central (LC), que representa donde la característica del proceso debería de caer



si no hay fuentes inusuales de variabilidad presente. La siguiente figura, muestra lo ya mencionado:



**Figura 1.1. Estructura de un gráfico de control**

*Fuente: Elaboración propia.*

Si el proceso se encuentra bajo control, la mayor cantidad de puntos se hallarán dentro de los límites, pero en cambio, cuando existen fuentes inusuales de variabilidad, la gráfica de control proyectará puntos fuera de los límites de control y así, una señal para realizar una investigación que descubra sus causas y se tomen medidas correctivas para eliminarlas (Montgomery, 2007).

Las gráficas de control son consideradas también como un contraste de hipótesis, establecida de la siguiente manera:

- $H_0$ : El proceso se encuentra bajo control
- $H_a$ : El proceso se encuentra fuera de control

En el cual, las cartas de control realizan este contraste repetidamente en todas las muestras que se toman a través del tiempo y así, rechazar o no la hipótesis nula ( $H_0$ ).

Como se mencionó en el apartado 1.6.1.1, se considerarán dos tipos de errores, que en el caso de gráficos de control son interpretados de la siguiente forma:

- Error Tipo I: Definir al proceso fuera de control cuando realmente no lo está.
- Error Tipo II: Definir al proceso bajo control cuando en la realidad sucede lo contrario.

Los cuales se encuentran relacionados de manera inversa ya que, si se reduce el riesgo de error tipo I, aumentará el error tipo II y viceversa (Wackerly, Mendenhall, & Scheaffer, 2010). Aquellos errores tendrán una probabilidad de ocurrencia  $\alpha$  y  $\beta$  respectivamente.

### **1.6.2.3. Construcción de gráficos de control**

Una gráfica de control sigue la evolución de varias características del 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, 2007).

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-Bustos, 2014).

Si la característica de calidad no es medible, se controla el número de defectos o unidades defectuosas por lote; en estos casos la función de probabilidad es Poisson o Binomial, en el cual se puede utilizar el teorema del límite central para hacer aproximaciones. En la siguiente tabla se presentan los parámetros más comunes a controlar y sus estimadores muestrales:

**Tabla 2. Parámetros más comunes a controlar y sus estimadores**

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

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

Para los gráficos de control de los cuatro primeros parámetros, se utilizan estadísticos cuya muestra se supone deben pertenecer a una población normalmente distribuida, en el cual el teorema de límite central es muy útil para estos casos (Montgomery, 2007).

Cuando se desea realizar un control de calidad por atributos, como se especifican en los cuatro últimos parámetros de la tabla, deben considerarse otras distribuciones para los estimadores. En el caso de la proporción de unidades disconformes de una muestra,  $\hat{p} = \frac{D}{n}$ , la distribución de probabilidad que sigue el estimador D, es Binomial; a diferencia de la cantidad de defectos por unidad de observación c, que tiene una función de probabilidad Poisson (García-Bustos, 2014).

El criterio más usado para establecer los límites de control en dichos gráficos, es aquel denominado 3 sigma, que establece que los límites deben encontrarse a tres desviaciones estándar arriba y abajo del valor esperado del estimador (Montgomery, 2007). En la siguiente tabla se presenta la forma de calcular el límite de control superior (LCS), inferior (LCI) y la línea central (LC) de los gráficos de control univariados más usuales:

**Tabla 3. Estructura de los gráficos de control univariados más usuales**

Gráfico de control	Estructura
Para media ( $\bar{X}$ ) con $n \leq 10$	$\begin{cases} LCS = \bar{\bar{X}} + A_2\bar{R} \\ LC = \bar{\bar{X}} \\ LCI = \bar{\bar{X}} - A_2\bar{R} \end{cases}$
Para media ( $\bar{X}$ ) con $n > 10$	$\begin{cases} LCS = \bar{\bar{X}} + A_3\bar{S} \\ LC = \bar{\bar{X}} \\ LCI = \bar{\bar{X}} - A_3\bar{S} \end{cases}$

Para el rango ( $R$ )	$\begin{cases} LCS = D_4\bar{R} \\ LC = \bar{R} \\ LCI = D_3\bar{R} \end{cases}$
Para la desviación estándar ( $S$ )	$\begin{cases} LCS = B_4\bar{S} \\ LC = \bar{S} \\ LCI = B_3\bar{S} \end{cases}$
Para la proporción de unidades defectuosas en la muestra ( $\hat{p}$ )	$\begin{cases} LCS = \bar{p} + 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \\ LC = \bar{p} \\ LCI = \bar{p} - 3\sqrt{\frac{\bar{p}(1-\bar{p})}{n}} \end{cases}$
Para el número de unidades defectuosas en la muestra ( $n\hat{p}$ )	$\begin{cases} LCS = n\bar{p} + 3\sqrt{n\bar{p}(1-\bar{p})} \\ LC = n\bar{p} \\ LCI = n\bar{p} - 3\sqrt{n\bar{p}(1-\bar{p})} \end{cases}$
Para el número de defectos por unidad de muestreo ( $c$ )	$\begin{cases} LCS = \bar{c} + 3\sqrt{\bar{c}} \\ LC = \bar{c} \\ LCI = \bar{c} - 3\sqrt{\bar{c}} \end{cases}$
Para la proporción de defectos en la muestra ( $u$ )	$\begin{cases} LCS = \bar{U} + 3\sqrt{\frac{\bar{U}}{n}} \\ LC = \bar{U} \\ LCI = \bar{U} - 3\sqrt{\frac{\bar{U}}{n}} \end{cases}$

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

Donde los coeficientes  $A_2, A_3, D_3, D_4, B_3$  y  $B_4$  serán calculados en función del tamaño muestral  $n$ , bajo el supuesto de normalidad ya mencionado. Existen tablas que nos especifican el valor de dichos coeficientes a través del tamaño muestral  $n$ .

Para este proyecto de investigación se plantearán nuevos esquemas de control multivariantes, aquellos que siguen el mismo planteamiento que los gráficos de control univariantes y serán detallados a profundidad en el capítulo siguiente.

#### 1.6.2.4. Medida de rendimiento ARL

Una medida de rendimiento para los gráficos de control es aquella denominada ARL (Average Run Length), aquella que establece el número promedio de muestras hasta que el proceso presenta una señal. Cabe indicar que, si las muestras son independientes, dicha medida de rendimiento es la media de una variable aleatoria geométrica con parámetro  $q$  (Montgomery, 2007), siendo ésta la probabilidad de que el gráfico muestre una señal; definiendo al ARL como  $1/q$ .

Para que un gráfico de control tenga un buen rendimiento, lo ideal es que el valor del ARL sea grande cuando el proceso se encuentra bajo control ( $ARL_0$ ), el cual puede ser calculado como  $1/\alpha$ , siendo  $\alpha$  la probabilidad de cometer el error tipo I. Además, se desea que el valor del ARL sea pequeño cuando el proceso se encuentra fuera de control ( $ARL_1$ ), en el cual, siendo  $\beta$  la probabilidad de cometer el error tipo II, la forma de calcular el  $ARL_1$  estará definida por  $1/(1 - \beta)$ . Al diseñar el gráfico de control, el investigador puede establecer un valor de  $\alpha$ , pero no de  $\beta$ , ya que éste último está asociado con el tamaño de muestra  $n$ ; por lo tanto, se deben fijar valores idóneos de  $\alpha$  y  $n$  para controlar los dos tipos de errores ya detallados, siendo  $\alpha$  y  $\beta$  los más pequeños posibles, pero no exactamente cero (Montgomery, 2007).

Entonces, se deduce que el diseño óptimo de un gráfico de control resulta aquel que logre valores grandes del  $ARL_0$  (establecido por el investigador, a través del valor de  $\alpha$ ) y valores pequeños del  $ARL_1$ , ya que así dicha herramienta detectará rápidamente los cambios inesperados en el proceso.

### 1.6.3. Gráfico de control T<sup>2</sup> de Hotelling

Mahalanobis (1936) introdujo una expresión matemática que mide la distancia entre dos variables aleatorias  $(X, Y)$  p-dimensionales, con igual función de distribución de probabilidades y matriz de varianzas y covarianzas  $\Sigma$ , definido de la siguiente manera:

$$d_m(X, Y) = \sqrt{(X - Y)^T \Sigma^{-1} (X - Y)}$$

**(1.1)**

Por lo tanto, sea  $X^T = (X_1, X_2, \dots, X_p)$  un conjunto de variables con vector de medias  $\mu^T = (\mu_1, \mu_2, \dots, \mu_p)$  y matriz de varianzas y covarianzas  $\Sigma_{p \times p}$  simétrica y definida positiva, el cuadrado de la distancia de Mahalanobis, para el conjunto de variables  $X$  estandarizado, será establecido por:

$$d_m^2(X, Y) = (X - \mu)^T \Sigma^{-1} (X - \mu)$$

**(1.2)**

Para lo cual, si  $X$  proviene de una distribución normal multivariante  $X \sim N_p(\mu, \Sigma)$  se define al estadístico  $\chi_p^2$  con una distribución Chi-cuadrada con p grados de libertad de la siguiente manera:

$$\chi_p^2 = n(\bar{X} - \mu)^T \Sigma^{-1} (\bar{X} - \mu)$$

**(1.3)**

Donde  $\bar{X}$  es el vector de medias muestral. Considerando que en la mayoría de las ocasiones no se conocen los parámetros de un proceso y hay que estimarlos, Harold Hotelling (1947) estableció un estadístico T<sup>2</sup>, a utilizarse en un gráfico de control para un conjunto de variables que siguen una distribución normal multivariada, definido por la fórmula matemática:

$$T^2 = n(\bar{X} - \hat{\mu})^T S^{-1} (\bar{X} - \hat{\mu})$$

**(1.4)**

Donde  $S$  y  $\hat{\mu}$  son los estimadores de la matriz de varianzas y covarianzas y del vector de medias, respectivamente. La estimación se la realiza con  $m$  muestras de tamaño  $n$  y si  $m$  tiende al infinito, la distribución del estadístico  $T^2$  converge a una distribución Chi-cuadrada con  $p$  grados de libertad (Seber, 1984).

Si se desea diseñar el gráfico  $T^2$ , sus límites de control serán definidos por  $LCS = \chi_{\alpha,p}^2$  y  $LCL = 0$ , donde  $\alpha$  es el nivel de significancia y  $p$  el número de variables de calidad monitorizadas.

En el caso de que el proceso de calidad, sufra un cambio en al menos una de las medias de sus variables, es decir, el vector  $\mu_1$  se aleja del vector de promedios  $\mu_0$  del estado bajo control en una distancia de Mahalanobis  $d$ , entonces el estadístico  $T^2$  tiene una distribución Chi-cuadrada no centralizada con  $p$  grados de libertad y con descentrado:

$$\lambda = nd^2 = n(\mu_1 - \mu_0)^T \Sigma^{-1} (\mu_1 - \mu_0)$$

**(1.5)**

Donde  $n$  es el número de observaciones o tamaño de muestra y  $\Sigma$  la matriz de varianzas y covarianzas de  $X$ , durante el proceso bajo control. Para este gráfico, el ARL es obtenido por la siguiente fórmula:

$$ARL = \frac{1}{1 - P(T^2 < LCS)}$$

**(1.6)**

Siendo LCS el límite de control superior, se asumirá que, si el estadístico cae exactamente en el límite, entonces existe una señal de que el proceso pudiera estar fuera de control (Hotelling, 1947).

Unos de los varios problemas de la gráfica de control  $T^2$  de Hotelling, es su insensibilidad a cambios pequeños en el vector de medias del



proceso o a medida que el número de variables  $p$ , se incrementa (Montgomery, 2007); es por ello, que se realizarán las comparaciones respectivas con los gráficos de control propuestos en esta investigación y ver si efectivamente se reducen dichos problemas, al existir un mejor rendimiento en el cálculo del ARL fuera de control.

#### **1.6.4. Algoritmos Genéticos (GA)**

Los algoritmos genéticos (Genetic Algorithm, en inglés) constituyen una técnica de optimización heurística que, para el caso de este proyecto de investigación, ayudará a encontrar los parámetros óptimos de los gráficos de control planteados en el mismo, detallando así, su base teórica.

Al hablar de algoritmos genéticos es necesario mencionar los estudios realizados por Charles Darwin y su teoría de evolución, ya que en su libro “El origen de las especies” en el año de 1859, propone que “las especies tienen su origen a través de selección natural, donde las razas más fuertes sobreviven respecto a otra más débiles y evolucionan sobre estas últimas” (Darwin, 2003). Por lo tanto, en la década de los 50's, algunos investigadores en el área de programación adoptaron esta ideología para construir nuevos algoritmos basados en reglas empíricas-rationales. Aquella forma de programación la denominaron “Computación evolutiva” que pretende emular las teorías Darwinistas para que, en una sociedad, inspirados en la evolución y la variación genética natural, la población pueda evolucionar hasta la resolución de un problema en general.

La técnica heurística GA, fue desarrollada en 1970 por John Henry Holland, quien creó algunos algoritmos genéticos para su implementación en inteligencias artificiales; siendo así el padre de dicha técnica. Los GA son métodos adaptivos que pueden usarse para resolver problemas de búsqueda y optimización, están basados

en el proceso genético de los organismos vivos, siendo capaces de ir creando soluciones que irán evolucionando hacia valores óptimos del problema planteado, donde ello dependerá en buena medida de una adecuada codificación de las mismas (Holland, 1992).

Se la considera como técnica robusta, y pueden tratar con éxito una gran variedad de problemas provenientes de diferentes áreas, incluyendo aquellos en los que otros métodos encuentran dificultades, siendo importante considerar lo siguiente (Moujahid, Inza, & Larrañaga):

- Se requerirán más iteraciones del algoritmo, si la función a optimizar tiene demasiados máximos/mínimos locales, para “asegurar” el máximo/mínimo global.
- Si la solución óptima de la función a optimizar, contiene varios puntos cercanos, solamente podemos “asegurar” que encontraremos uno de ellos y no necesariamente el óptimo.

En la siguiente tabla, se presentan los conceptos de los elementos básicos utilizados en los algoritmos genéticos:

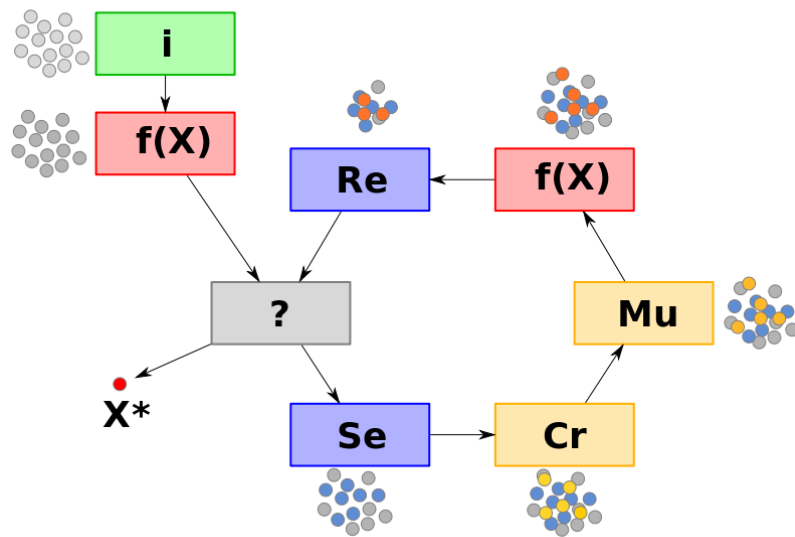
**Tabla 4. Conceptos básicos utilizados en algoritmos genéticos (GA).**

Elemento	Concepto
Individuo	Posible solución al problema establecido.
Población	Conjunto de posibles soluciones/individuos.
Gen	Representación de un parámetro específico del individuo.

Cromosoma	Agrupación de genes de un individuo en el cuál, la forma más utilizada para su representación es la codificación binaria, siendo más efectiva que otros métodos de codificación existentes.
-----------	---

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

A continuación, mostraremos a detalle las fases que contemplan la implementación de algoritmos genéticos:



**Figura 1.2. Fases en la ejecución de un algoritmo genético**

*Fuente: Wikipedia. Algoritmo Genético, 2017.*

1. Iniciación (i): Se crean conjuntos de cromosomas (individuos aleatorios), los cuales podrían dar solución al problema.

Mientras el número de evaluaciones sea menor al máximo de evaluaciones establecido en el algoritmo o cuando no haya ningún tipo de cambio en el grupo de individuos, se estará realizando lo siguiente:

2. Evaluación  $f(X)$ : Se aplicará una función de aptitud, a cada uno de los cromosomas, que certifique cuan buena es la solución que aporta el individuo.
3. Selección (Se): Aquellos individuos con mejores resultados, aplicando la función de aptitud de la fase anterior, tendrán mayor probabilidad de ser seleccionados. Existen varias formas de selección, entre las más comunes: Por torneo, por ruleta o por rango.
4. Cruzamiento (Cr): Representa la reproducción de los individuos, siendo éste el principal operador a nivel genético, donde utilizando dos cromosomas se generan dos descendientes que combinan las características de los cromosomas padres. Entre las formas de cruzamiento, tenemos: Un punto, dos puntos, uniforme y aritmético.
5. Mutación (Mu): Se realizarán modificaciones al azar del cromosoma de los individuos, lo que ayudará a considerar nuevas opciones que permitan analizar esos espacios que no están cubiertos en la población estudiada por el momento. Dicha operación se realiza a través de una probabilidad de mutación, que suele ser baja para evitar una búsqueda aleatoria en los GA.
6. Reemplazo (Re): Una vez realizadas las fases anteriores, se seleccionarán a los mejores individuos del grupo analizado, para generar al conjunto de cromosomas de la siguiente generación.

Cabe mencionar que, los algoritmos genéticos (GA) pueden presentarse de distintas maneras, ya que existen variaciones en las operaciones de genética en los individuos, formando así, una población diferente por cada variación (Holland, 1992).

Para el diseño de los gráficos de control planteados en este proyecto de investigación y detallados en el siguiente capítulo, se utilizaron

algoritmos genéticos para encontrar los parámetros óptimos que minimicen el ARL fuera de control, en el cuál los métodos de selección que se probaron fueron los de torneo y ruleta junto al tipo de cruzamiento de un punto, dos puntos y uniforme, con una probabilidad de ocurrencia del 0.01, 0.05, 0.10 y 0.15.

Detallando un poco los métodos de selección utilizados, tenemos aquel denominado selección por torneo, que escoge al azar un determinado número de individuos para reproducir al más apto; y aquel denominado selección por ruleta, que consiste en crear una ruleta con los cromosomas de cada generación donde tendrán mayor probabilidad de ser seleccionados, aquellos que tengan una mejor puntuación. Las formas de cruzamiento de un punto o dos, cortan los dos cromosomas a reproducirse en dichos puntos específicos, donde para el caso del cruce en un punto, el nuevo individuo tendrá como cromosoma una parte de un progenitor y otra parte del segundo, diferente para el caso de dos puntos ya que el individuo a formarse tendrá dos partes del primer progenitor (su cromosoma inicial y final), y una parte del segundo progenitor (la parte media de su cromosoma). Para el cruzamiento de forma uniforme, los padres aportan a su descendiente sus genes de manera aleatoria (Holland, 1992). Las funciones de aptitud utilizadas, son aquellas recomendadas por Aparisi y García-Díaz (2007), ya que aplican el concepto de diseño óptimo sobre un punto, que consiste en especificar el  $ARL_0$  deseado a un vector de cambios  $d^*$  que se desea detectar.

Por lo tanto, se obtiene la combinación de parámetros del gráfico que obtiene el mínimo ARL fuera de control  $ARL(d=d^*)$ , restringido a un  $ARL_0$  deseado. Dichas funciones, se detallan a continuación:

$$1. \quad \max - w_1 |ARL_0 - ARL(d = 0)| - w_2 ARL(d = d^*)$$

$$\text{si } (ARL_0 - ARL(d = 0)) \leq 0$$

**(1.7)**

$$2. \quad \max - 2w_1|ARL_0 - ARL(d = 0)| - w_2ARL(d = d^*)$$

$$\text{si } (ARL_0 - ARL(d = 0)) > 0$$

**(1.8)**

En el cual, el  $ARL_0$  representa el ARL deseado en el proceso y el  $ARL(d=0)$  detalla el ARL bajo control que se obtiene con los parámetros del gráfico que lleva cada uno de los individuos a evaluar. El valor de ARL para detectar el cambio  $d^*$  que el gráfico de control debe minimizar, está detallado por  $ARL(d=d^*)$ , donde, los coeficientes  $w_1$  y  $w_2$  representan las ponderaciones de cada objetivo a optimizar en cada ecuación, aquellos que han sido establecidos en el algoritmo por 1 y 5, respectivamente; tal como lo recomienda Aparisi y García-Díaz en trabajos anteriores.

## **1.7. Revisión Bibliográfica**

A continuación, se resumirán algunos artículos o investigaciones realizados en el área del control estadístico a través de técnicas o gráficas de control multivariantes.

### **1.7.1. Selección de variables fuera de control con el procedimiento de calidad multivariante $T^2$**

“Aunque la calidad se mide a menudo por el nivel de correlación de un conjunto de varias variables aleatorias, es común encontrar en la práctica, las decisiones de control basadas en los gráficos individuales de Shewhart” menciona Murphy en este estudio.

Hotelling (1947) propuso el uso de un gráfico de control multivariado basado en su estadístico  $T^2$ , suponiendo que la distribución subyacente de las variables aleatorias es normal multivariante. Sin embargo, el autor de ésta investigación menciona que, la falta de un método simple de selección de variables "fuera de control" ocasiona que el uso rutinario de las gráficas  $T^2$  no sea práctico.

En este artículo, se proporciona una prueba sencilla para seleccionar las variables fuera de control y realizar una interpretación de los valores del estadístico  $T^2$ . Las ideas están motivadas para considerar el control de calidad como un ejemplo de Análisis Discriminante. Dicho enfoque se ilustra en un ejemplo bivariado y luego en dos métodos de selección con cuatro variables de respuesta, recomendando que las cartas individuales no deben usarse para propósitos de control o incluso para escoger variables fuera de control (Murphy, 1987).

### **1.7.2. Control estadístico de procesos multivariados**

MacGregor y Kourti detallan que el avance tecnológico capta un gran número de información acerca de las variables que tienen los principales procesos industriales y los análisis con métodos estadísticos multivariados, el monitoreo y el diagnóstico del funcionamiento operativo del proceso mencionado, han recibido una creciente atención. Las extensiones de los gráficos de control univariados tipo Shewhart y los gráficos CUSUM y EWMA a situaciones de control de calidad, se basan por lo general en el estadístico  $T^2$  de Hotelling.

Los enfoques para el control estadístico multivariado, a mediados de la década de los 90's, utilizaban las variables de calidad más importantes del proceso y la información relevante que conlleva al producto final. Estos métodos se basan en el análisis de componentes principales (PCA) y mínimos cuadrados parciales (PLS).

En este trabajo se ofrece una visión general de su utilización, para el proceso de control estadístico de variables continuas y multivariadas donde a través de un ejemplo de aplicación, se ilustra cómo debe ser su aplicación para analizar las operaciones en una planta de procesamiento de minerales (MacGregor & Kourti, 1995).

### **1.7.3. Monitoreo y control en un proceso normal multivariado**

Zamarrón et al. establecen una metodología para controlar y monitorear procesos normales multivariados a través de la gráfica de control  $T^2$  de Hotelling, que detecta las señales fuera de control y el análisis de componentes principales que ayudarán a identificar las variables que ocasionarán dicha señal y la estructura dimensional del producto.

El trabajo se enfoca en proporcionar a los responsables del control estadístico de procesos información oportuna y adecuada para la correcta toma de decisiones en el control y monitoreo de procesos normales multivariados. Ambas metodologías proporcionan datos valiosos para decidir en forma objetiva y práctica cuándo y dónde ajustarlo. (Zamarrón, Prado & Luis, 2012).

### **1.7.4. Combinación lineal óptima de variables Poisson para el control estadístico multivariante de procesos.**

“La distribución de Poisson es adecuada para modelar la frecuencia del número de defectos en un proceso de producción y útil para desarrollar su gráfico de control, sin embargo, hay casos en que su distribución no es adecuada” detallan Epprecht, García-Bustos y Aparisi en una investigación realizada en el año 2013.

En este artículo, se analiza el seguimiento de  $p$  características de calidad tipo Poisson de forma simultánea, desarrollando un nuevo gráfico de control multivariado basado en la combinación lineal de dichas variables, denominado gráfico de control LCP. Los coeficientes de esta combinación lineal se optimizan para obtener un ARL bajo control deseado y minimizar el ARL fuera de control.

Para facilitar el uso de este nuevo diagrama, la optimización se realizó empleando un software en Windows © fácil de usar, que compara el rendimiento entre este y otros esquemas multivariados basados en el



monitoreo de un conjunto de variables Poisson, tales como un gráfico de control sobre la suma de las variables (gráfico MP), una carta de control para máximos (gráfico MX) y un conjunto optimizado de gráficos de Poisson univariantes (esquema múltiple).

Concluyen en este artículo que el gráfico de control LCP diseñado por los autores, muestra un rendimiento muy bueno. En primer lugar, se logra un ARL bajo control deseado, que es una ventaja sobre el resto de gráficos, que no pueden alcanzar este valor requerido, porque sus límites sólo pueden tomar valores enteros. En segundo lugar, se observó que, en la mayoría de los casos, este esquema detecta mucho más rápido los cambios en el proceso que el resto de los gráficos (Epprecht, García-Bustos & Aparisi, 2013).

# CAPÍTULO 2

## 2. METODOLOGÍA DEL DISEÑO

Este capítulo describe la base teórica de los gráficos de control multivariados propuestos, basados en combinaciones lineales de variables normales y de componentes principales; además se expone brevemente el problema de optimización planteado para encontrar mediante GA los parámetros óptimos del gráfico de control LCN.

Como se mencionó en la introducción de esta investigación, para medir el rendimiento de estos esquemas multivariados se utiliza como medida de rendimiento la longitud promedio de corrida (ARL), definida como el número promedio de puntos en el gráfico de control hasta que presente una señal, ya sea, fuera de los límites de control o simplemente en los límites de control. Por lo general, cuando no se presentan cambios en los parámetros del proceso, se requiere que el ARL bajo control tome un gran valor, pero en cambio, cuando existen dichos cambios (proceso fuera de control) se desea que el gráfico los detecte rápidamente, requiriendo así que los valores del  $ARL_1$  sean los más pequeños posibles. Para los gráficos analizados en este trabajo, donde los puntos a trazar son independientes, el número promedio de muestras hasta que el gráfico presente una señal será la media de una variable aleatoria geométrica, definida como  $ARL = 1 / P(\text{señal})$ .

### 2.1. Gráfico de control basado en la combinación lineal de variables normales correlacionadas (LCN control chart).

Sean  $X_i$  variables aleatorias normales correlacionadas provenientes de una distribución normal multivariante  $N(\mu, \Sigma)$ , con  $i = 1, 2, \dots, p$ , y a través de un proceso de estandarización, se define a  $Z_i$  con la distribución multivariante  $N(\vec{0}, R)$  siendo  $R$  la matriz de correlación del conjunto de

variables  $X_i$ . El estadístico utilizado para este gráfico de control es definido por  $LCN = a_1Z_1 + a_2Z_2 + \dots + a_pZ_p$ , donde  $a_i \in [-1,1]$ , y se encuentra normalmente distribuido (Rencher, 2003) de la siguiente manera:

$$LCN \sim N\left(0, \sum_{i=1}^p a_i^2 Var(Z_i) + 2 \sum_{i=1}^p \sum_{i<j} a_i a_j Corr(Z_i, Z_j)\right) \quad (2.1)$$

Para lo cual, estableciendo sus parámetros de manera matricial, se define la distribución del estadístico LCN como:

$$LCN \sim N(0, a^T R a)$$

(2.2)

Donde:

$$a^T: (a_1, a_2, \dots, a_p)^T \quad \text{y} \quad R: \begin{pmatrix} 1 & \dots & \rho_{1p} \\ \vdots & \ddots & \vdots \\ \rho_{p1} & \dots & 1 \end{pmatrix}$$

(2.3)

(2.4)

Dado que una combinación lineal de variables normales puede tomar valores negativos, el gráfico LCN requiere de dos límites de control, uno superior y uno inferior. Por lo tanto, la expresión para el cálculo del ARL será llevada a cabo de la siguiente manera:

$$ARL = \frac{1}{1 - P(LCI < LCN < LCS)}$$

(2.5)

Donde LCI es el límite de control inferior y LCS es el límite de control superior para el gráfico LCN. En el caso que se trabaje con una muestra de tamaño  $n$  mayor a uno, el estadístico a usar es la media muestral de las combinaciones lineales con una distribución:

$$N(0, (a^T R a)/n)$$

(2.6)

## 2.2. Gráfico de control basado en la combinación lineal de componentes principales (LCPC control chart).

Como el objetivo de las componentes principales es reducir la dimensionalidad de las variables con las que se trabaja ( $p$  variables) a un subconjunto  $q$  menor ( $q \leq p$ ), con la condición de perder una pequeña parte de la variabilidad total del proceso, se decidió aplicar dicha metodología para definir un estadístico basado en la combinación lineal de al menos las dos primeras componentes principales que expliquen como mínimo el 80% de la variabilidad de las variables originales.

Aquella metodología consiste en representar adecuadamente la información de  $n$  observaciones de  $p$  variables observables, a través de un número menor de variables independientes  $q$  no observables, obtenidas a través de combinaciones lineales de las originales, representando dichas variables latentes como  $C_i$ , de la siguiente manera:

$$\begin{cases} C_1 = \alpha_{11}x_1 + \alpha_{12}x_2 + \dots + \alpha_{1p}x_p \\ C_2 = \alpha_{21}x_1 + \alpha_{22}x_2 + \dots + \alpha_{2p}x_p \\ \vdots \\ C_p = \alpha_{p1}x_1 + \alpha_{p2}x_2 + \dots + \alpha_{pp}x_p \end{cases}$$

(2.7)

Donde  $\alpha_{ij}$  son los coeficientes de cada combinación lineal y  $x_p$  representan el conjunto de variables del proceso a controlar. Para lo cual, la media de cada componente será representada por  $E(C_i) = b_i^t \mu$  y su varianza por  $V(C_i) = b_i^t \Sigma b_i$ , tal que  $b_i$  es el vector de coeficientes de las combinaciones lineales o componentes  $C_i$  y  $\Sigma$  representa la matriz de varianzas y covarianzas de  $x$ . Encontrar los coeficientes  $\alpha_{ij}$  es simple, ya que si  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p \geq 0$  son los valores propios de  $\Sigma$  ordenados de mayor a menor, los vectores  $b_i$  son simplemente el  $i$ -ésimo vector propio asociado con el valor propio  $\lambda_i$ , considerando que la primera componente es aquella asociada al más grande valor propio, la segunda componente tiene asociado el segundo valor propio más alto y así,

sucesivamente. Básicamente si  $A$  es la matriz cuyas columnas son los vectores propios, entonces:

$$A^T \Sigma A = \Lambda$$

(2.8)

Donde  $\Lambda$  es una matriz diagonal de dimensión  $p \times p$  tal que, en su diagonal se encuentran ubicados todos los posibles valores propios  $\lambda_i$ , que representan a su vez la varianza de cada componente  $C_i$ . Por consiguiente, la proporción de variabilidad en los datos originales explicado por cada componente, estará dada por la siguiente relación:

$$\frac{\lambda_i}{\lambda_1 + \lambda_2 + \dots + \lambda_p}$$

(2.9)

Por lo tanto, la variabilidad explicada a través de  $q$  componentes principales se calcula a través de la suma de los valores propios de dicha cantidad de componentes, comparado con la suma de todos los  $p$  valores propios (Rencher, 2003; Montgomery, 2007). Como el enfoque de esta investigación es trabajar con variables estandarizadas, la matriz de varianzas y covarianzas  $\Sigma$  cambia por la matriz de correlación  $R$  para los cálculos respectivos; la razón de ello es que “a menudo los procesos presentan muchas variables con escalas diferentes y como consecuencia una variable puede parecer que contribuye mucho a la variabilidad total del sistema, simplemente porque su escala de medida tiene magnitudes mayores que las otras variables” (Montgomery, 2007). Una vez explicado cómo se lleva el cálculo de las componentes principales del proceso, se procederá a definir el estadístico LCPC de la siguiente manera:

$$LCPC = e_1 C_1 + e_2 C_2 + \dots + e_q C_q$$

(2.10)

Donde  $e_i \in [-1,1]$  representan los coeficientes encontrados para al menos las dos primeras  $C_q$  componentes principales que representen al

menos el 80% de la variabilidad del proceso. La distribución del estadístico LCPC, basado en la combinación lineal de componentes principales, se define como una normal univariada (Wackerly, Mendenhall, & Scheaffer, 2010), de la siguiente manera:

$$LCPC \sim N\left(0, \sum_{i=1}^q e_i^2 \lambda_i\right) \text{ tal que } i = 1, 2, \dots, q \text{ y } 2 \leq q \leq p$$

**(2.11)**

Como en el caso anterior, de la combinación lineal de variables normales, el estadístico LCPC puede tomar valores negativos y, por lo tanto, su gráfico de control estará definido por un límite inferior y otro superior; estableciendo el cálculo del ARL de igual manera:

$$ARL = \frac{1}{1 - P(LCI < LCPC < LCS)}$$

**(2.12)**

En el caso que se trabaje con una muestra de tamaño  $n$  mayor a 1, el estadístico a usar es la media muestral de las combinaciones lineales de las componentes principales con una distribución:

$$N\left(0, \sum_{i=1}^q (e_i^2 \lambda_i) / n\right)$$

**(2.13)**

### **2.3. Problema de optimización para el gráfico de control LCN y LCPC**

Para la optimización de los parámetros de los gráficos de control propuestos, se utilizaron Algoritmos Genéticos (GA). Dicha técnica de optimización heurística, ha sido programada en Visual Basic Community©, con la finalidad de obtener una interfaz que nos ayude a visualizar los diferentes resultados obtenidos con la técnica. Como referencia consideramos los trabajos realizados por Aparisi et al. (Aparisi

& de Luna, 2009), Epprecht et al. (Epprecht, García-Bustos, & Aparisi, 2013) y García-Bustos et al. (García-Bustos, Aparisi, & Epprecht, Optimal EWMA of Linear Combination of Poisson Variables for Multivariate Statistical Process Control, 2014). La definición formal del problema de optimización sería el siguiente:

Dado:

ARL Bajo control:  $ARL_0$

Medias bajo control:  $\mu_i \quad i = 1, 2, \dots, p$

Los valores de la matriz de correlación  $R: \begin{pmatrix} 1 & \cdots & \rho_{1p} \\ \vdots & \ddots & \vdots \\ \rho_{p1} & \cdots & 1 \end{pmatrix}$

Cambios para el cual minimizar el ARL's:  $d^* = (d_1, d_2, d_3, \dots, d_p)$

Donde  $d_i$ : Cambio en sigma unidades para la media  $\mu_i$

Encontrar:

Límite de control superior e inferior para las p variables analizadas:  
 $LCS$  y  $LCL$

Los parámetros  $a_i \in [-1, 1]$ ,  $i = 1, 2, \dots, p$  para la combinación lineal  $LCN = \sum_{i=1}^p a_i Z_i$  o los parámetros  $e_j \in [-1, 1]$ ,  $j = 1, 2, \dots, q$  para la combinación lineal  $LCPC = \sum_{j=1}^q e_j C_j$

Minimizar:  $ARL(d = d^*) = ARL_1$

Sujeto a  $ARL(d = 0) = ARL_0$

# CAPÍTULO 3

## 3. RESULTADOS

Este capítulo presenta la utilización de los gráficos de control propuestos, en un ejemplo de aplicación que se ajuste a las necesidades de los gráficos LCN y LCPC y así, analizar rendimientos en comparación con el gráfico T<sup>2</sup> de Hotelling. Para exponer estos resultados, se ha desarrollado en Microsoft Visual Basic Community© enlazado con el software libre R (R Core), un programa que permita optimizar con Algoritmos Genéticos (GA) los parámetros de los gráficos ya mencionados.

### 3.1. Software y ejemplo de aplicación

“La necesidad de los gráficos de control en un proceso de manufactura es una estrategia importante para mejorar la calidad de producción” (Rojas, Hernández, & Ruiz, 2014), por ello se ha tomado como ejemplo el enfoque de estudio utilizado en la investigación realizada por Rojas, Hernández y Ruiz en un proceso de producción de galletas; en el cuál mencionan que las variables críticas que intervienen dentro de proceso de horneado de las mismas pueden ser: el peso y volumen de la galleta horneada, y el peso crudo maquinado, de las cuales sus unidades de medidas estarán establecidas en gramos (g) y centímetros cúbicos (cm<sup>3</sup>) respectivamente. Por ende, estas variables deben cumplir con las especificaciones descritas por el departamento de calidad de las industrias que realizan este tipo de procesos ya que, “Un incremento o disminución en el peso o volumen de la galleta, dará como resultado entre otras cosas un mal sellado en los paquetes que la contienen” (Rojas, Hernández, & Ruiz, 2014).

Tendrá sentido que dichas variables contengan una relación positiva ya que, al controlar el volumen y peso de la galleta en el proceso de



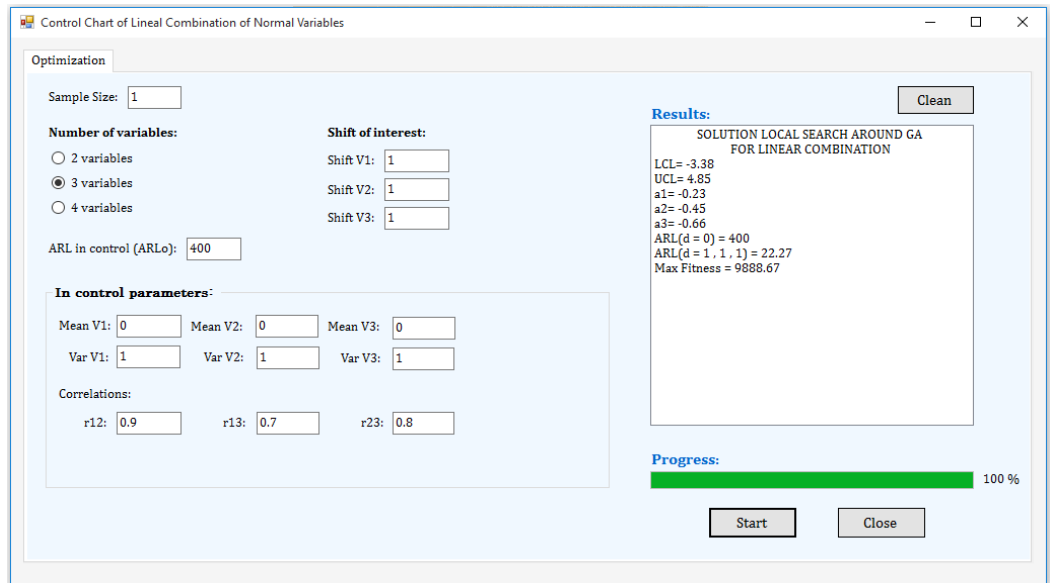
horneado, a medida que la galleta contenga mayor masa se obtendrá un mayor peso de las mismas; existiendo la misma correspondencia al medir el peso antes y después del horneado, por lo tanto, para el desarrollo del ejemplo de aplicación y como requisito del gráfico de control LCN se considera la siguiente matriz de correlación  $R$  para estas variables:

$$\begin{array}{r}
 \text{P. Crudo} \\
 \text{P. Horneado} \\
 \text{Volumen}
 \end{array}
 \begin{array}{c}
 \text{P. Crudo} \\
 \text{P. Horneado} \\
 \text{Volumen}
 \end{array}
 \begin{array}{c}
 \text{P. Horneado} \\
 \text{Volumen}
 \end{array}
 \begin{array}{c}
 \text{Volumen}
 \end{array}
 \left( \begin{array}{ccc}
 1.00 & 0.90 & 0.70 \\
 0.90 & 1.00 & 0.80 \\
 0.70 & 0.80 & 1.00
 \end{array} \right)$$

(3.1)

Se desea entonces obtener los parámetros óptimos de la carta de control LCN para detectar rápidamente los cambios en  $\mu_1$ ,  $\mu_2$  y  $\mu_3$  en una desviación estándar, restringidos a un  $ARL_0 = 400$ . Cabe recordar, que trabajaremos con el supuesto de que las variables se encuentran estandarizadas, por lo que el vector de medias cuando el proceso está bajo control es el vector  $\vec{0}$  con la matriz de correlación mencionada.

A continuación, se detallan las consideraciones que el usuario debe tomar para el correcto funcionamiento del software desarrollado: Primero, se debe seleccionar el número de variables a monitorear en la sección "Number of variables"; luego, establecer en "Sample Size" el número de muestras por lote a analizar; posteriormente introducir las medias, varianzas, correlaciones y los respectivos cambios en las medias de las características de calidad que desee controlar, junto al valor deseado del ARL bajo control ( $ARL_0$ ). Terminando con la introducción de los datos necesarios, se presiona el botón "Start" para obtener los siguientes resultados: El límite de control superior e inferior, los pesos óptimos para el estadístico LCN, el  $ARL_0$  y el ARL fuera de control ( $ARL_1$ ) minimizado para detectar rápidamente los cambios deseados. La siguiente figura muestra la salida del software:

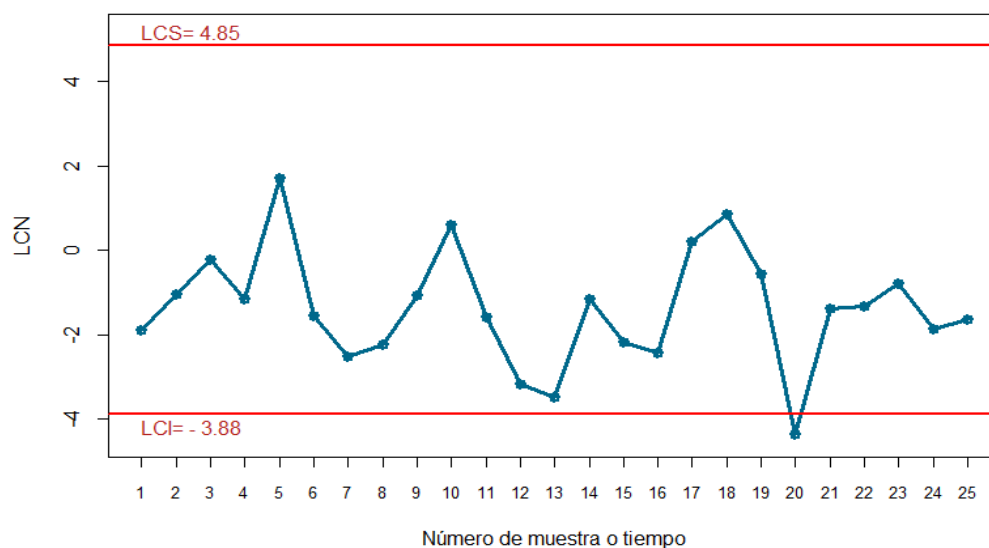


**Figura 3.1. Salida del software para el gráfico de control LCN optimizado para el ejemplo de aplicación propuesto.**

*Fuente: Elaboración Propia.*

Considerando los parámetros optimizados para el ejemplo de las variables críticas en el proceso de horneado de galletas, el gráfico de control estará establecido por el estadístico  $LCN = -0.23 Z_{P.Crudo} - 0.45 Z_{P.Horneado} - 0.66 Z_{Volumen}$ , tal que  $Z$  son las variables de calidad originales estandarizadas, donde los límites de control superior e inferior tomarán respectivamente un valor de 4.85 y -3.88. Dicho gráfico bajo control, mostrará en promedio una señal de falsa alarma en la muestra número 400, para lo cual, cuando existan incrementos de una desviación estándar en cada una de las medias de las variables analizadas, este cambio será detectado en promedio luego de que se hayan tomado 22 muestras de tamaño 1.

La siguiente figura ilustra el gráfico LCN cuando el proceso se encuentre fuera de control, con cambios de una desviación estándar en cada una de sus medias:



**Figura 3.2. Gráfico de control LCN para un proceso de horneado fuera de control, con datos simulados.**

*Fuente: Elaboración Propia.*

Dicha ilustración ha sido diseñada simulando 25 muestras con características de calidad mencionadas para el proceso de horneado de galletas (Véase Apéndice A) aquellas que, junto a su matriz de correlación, determinaron el estadístico LCN para cada una de las muestras, los cuales son esquematizados en los puntos de la gráfica. Se observa, que en la muestra 20 se encuentra una señal fuera de control, donde como se mencionó en el Capítulo 1, se debe parar la producción, realizar una investigación que descubra sus causas y tomar medidas correctivas para eliminarlas (Montgomery, 2007).

Por otro lado, si el usuario desea obtener los parámetros del gráfico de control LCPC, debe calcular primeramente los vectores y valores propios de la matriz de correlación. Para el caso del ejemplo de aplicación, los valores y vectores propios que se obtuvieron en el software R fueron los siguientes:

$$\lambda_1 = 2.6027$$

**(3.1)**

$$\lambda_2 = 0.3140$$

**(3.2)**

$$\lambda_3 = 0.0831$$

**(3.3)**

$$b_1 = (-0.58, -0.60, -0.55)$$

**(3.4)**

$$b_2 = (-0.57, -0.19, 0.80)$$

**(3.5)**

$$b_3 = (0.59, -0.78, 0.23)$$

**(3.6)**

Por lo tanto, las componentes principales serían:

$$C_1 = -0.58 Z_{P.Crudo} - 0.60 Z_{P.Horneado} - 0.55 Z_{Volumen}$$

**(3.7)**

$$C_2 = -0.57 Z_{P.Crudo} - 0.195 Z_{P.Horneado} + 0.80 Z_{Volumen}$$

**(3.8)**

$$C_3 = 0.59 Z_{P.Crudo} - 0.78 Z_{P.Horneado} + 0.23 Z_{Volumen}$$

**(3.9)**

Al ser los valores propios las varianzas de los componentes principales, se puede apreciar que tan solo la primera componente explica el 86.76% de la variabilidad total, donde para efectos prácticos se trabajarán con las dos primeras componentes principales, explicando así el 97.22%. Como se mencionó en el Capítulo 2, dichas variables latentes no están correlacionadas y al ser combinaciones lineales de variables estandarizadas, sus medias tomarán un valor de cero, donde los cambios respectivos  $d^* = (1,1,1)$  en las medias de las variables originales tomarán un incremento o decremento en las medias de las dos primeras componentes a analizar, de la siguiente manera:

$$E(C_1) = -0.58(1) - 0.60(1) - 0.55(1) = -1.73$$

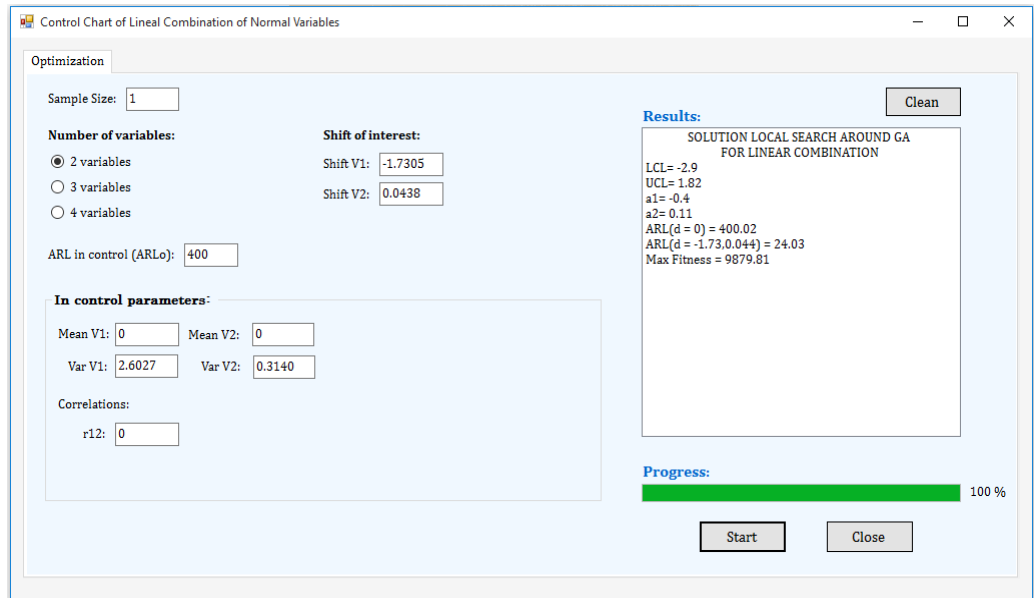
**(3.10)**

$$E(C_2) = -0.57(1) - 0.19(1) + 0.80(1) = 0.04$$

**(3.11)**

Dichas medias, varianzas, correlaciones y cambios en las medias a optimizar, de las componentes principales a analizar, deben ser

ingresadas al software diseñado junto al valor deseado del ARL bajo control ( $ARL_0$ ); para lo cual, al momento de presionar el botón “Start” se presentarán los parámetros óptimos para el gráfico de control LCPC, tal como se ilustra en la Figura 3.3:

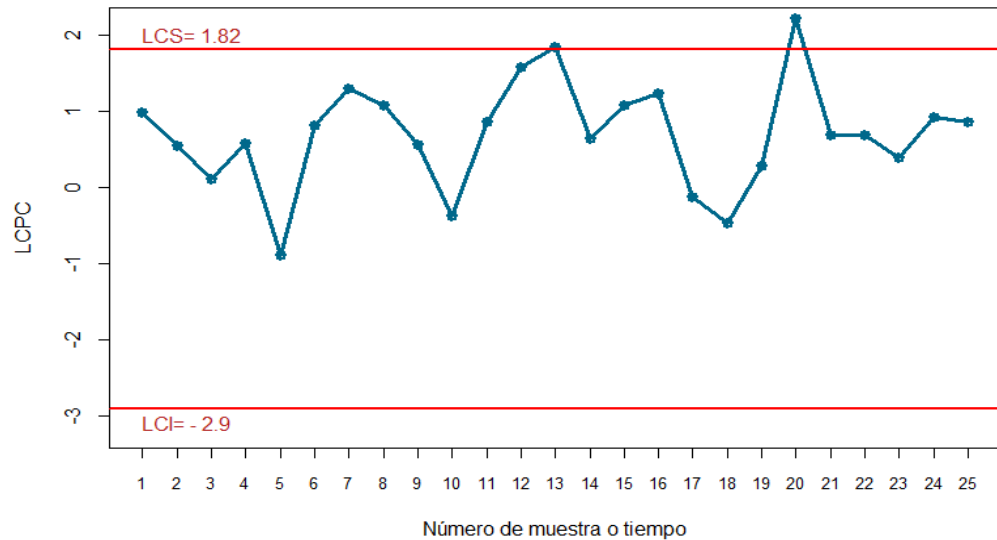


**Figura 3.3. Salida del software para el gráfico de control LCPC optimizado para el ejemplo de aplicación propuesto.**

*Fuente: Elaboración Propia.*

Por lo tanto, el gráfico de control LCPC para el proceso de horneado de galletas, estará establecido por el límite de control superior e inferior con un valor de 1.82 y -2.9, respectivamente; además, del estadístico  $LCPC = -0.4 C_1 + 0.11 C_2$ . Al igual que el gráfico LCN, el gráfico bajo control LCPC mostrará en promedio una señal de falsa alarma en la muestra número 400, para lo cual, cuando existan incrementos de una desviación estándar en cada una de las medias de las variables originales, este cambio será detectado en promedio luego de que se hayan tomado 24 muestras de tamaño 1, rendimiento muy similar al obtenido con el gráfico LCN puesto que, se está trabajando con las componentes que retienen un 97% de la variabilidad total.

La Figura 3.4, ilustra a través de los datos simulados anteriormente (Véase Apéndice B), lo mencionado al final del párrafo anterior:



**Figura 3.4. Gráfico de control LCPC para un proceso de horneado fuera de control, con datos simulados.**

*Fuente: Elaboración Propia.*

Observamos un comportamiento muy similar al gráfico de control LCN, en el cual la presente carta encuentra una señal fuera de control en la vigésima y décimo tercera muestra, realizando en cada uno de estos tiempos el procedimiento mencionado anteriormente.

Si se compararan estos gráficos con el  $T^2$  de Hotelling; el vector de cambios en desviaciones estándar  $d^* = (1, 1, 1)$  equivale a una distancia de Mahalanobis de 1.085 y con un valor de  $ARL_0=400$  entonces, el  $ARL_1$  para el gráfico  $T^2$ , obtenido a través del software R con la metodología descrita en el Capítulo 1, es de 76.30. Por consiguiente, el  $ARL_1$  de las cartas de control LCN y LCPC presentan una reducción del 70.84% y 68.51% en relación a la carta de control  $T^2$  respectivamente. Por lo tanto, las gráficas diseñadas en esta investigación presentan mejores rendimientos para controlar aquellas variables críticas en el proceso de manufactura, mencionado.

### 3.2. Comparación de rendimientos

Para este análisis se han propuesto tres escenarios para  $p=2$ ,  $p=3$  y  $p=4$  variables, considerando diversas matrices de correlaciones, valores de  $ARL_0$  y cambios en las medias de las variables originales de hasta dos desviaciones estándar; especificando a su vez, que en todos los casos las variables de calidad analizadas, son normales estándar. También, se consideró como tamaño muestral el valor de uno, aunque otros tamaños mayores fueron evaluados, obteniéndose resultados similares a los presentados en este apartado.

El programa desarrollado y explicado en el apartado anterior, permite obtener los valores de  $ARL_1$  para los gráficos de control propuestos en este proyecto, mientras que el  $ARL_1$  para el gráfico de control  $T^2$  fue obtenido mediante el software R con el marco teórico expuesto en el Capítulo 1. Los escenarios analizados se presentan en la Tabla 5.

**Tabla 5. Escenarios analizados en la comparación de rendimientos de los gráficos de control diseñados.**

Número de variables	Escenarios		
p=2	$R = \begin{pmatrix} 1 & 0.80 \\ 0.80 & 1 \end{pmatrix}$ $ARL_0 = 300$ Escenario (a)	$R = \begin{pmatrix} 1 & 0.65 \\ 0.65 & 1 \end{pmatrix}$ $ARL_0 = 500$ Escenario (b)	$R = \begin{pmatrix} 1 & 0.30 \\ 0.30 & 1 \end{pmatrix}$ $ARL_0 = 400$ Escenario (c)
p=3	$R = \begin{pmatrix} 1 & 0.85 & 0.5 \\ 0.85 & 1 & 0.6 \\ 0.5 & 0.6 & 1 \end{pmatrix}$ $ARL_0 = 400$ Escenario (d)	$R = \begin{pmatrix} 1 & 0.38 & 0.20 \\ 0.38 & 1 & 0.75 \\ 0.20 & 0.75 & 1 \end{pmatrix}$ $ARL_0 = 600$ Escenario (e)	$R = \begin{pmatrix} 1 & 0.22 & 0.45 \\ 0.22 & 1 & 0.18 \\ 0.45 & 0.18 & 1 \end{pmatrix}$ $ARL_0 = 800$ Escenario (f)
p=4	$R = \begin{pmatrix} 1 & 0.25 & 0.30 & 0.44 \\ 0.25 & 1 & 0.15 & 0.50 \\ 0.3 & 0.15 & 1 & 0.20 \\ 0.44 & 0.50 & 0.20 & 1 \end{pmatrix}$ $ARL_0 = 600$ Escenario (g)	$R = \begin{pmatrix} 1 & 0.70 & 0.50 & 0.45 \\ 0.70 & 1 & 0.38 & 0.35 \\ 0.50 & 0.38 & 1 & 0.66 \\ 0.45 & 0.35 & 0.66 & 1 \end{pmatrix}$ $ARL_0 = 800$ Escenario (h)	$R = \begin{pmatrix} 1 & 0.90 & 0.85 & 0.60 \\ 0.90 & 1 & 0.80 & 0.70 \\ 0.85 & 0.80 & 1 & 0.55 \\ 0.60 & 0.70 & 0.55 & 1 \end{pmatrix}$ $ARL_0 = 1000$ Escenario (i)

Fuente: Elaboración propia

Para  $p=2$  (Véase Apéndice C), en los distintos niveles de correlación, se observa que, para todos los cambios de las medias analizadas, el  $ARL_1$  obtenido con el gráfico de control LCN es menor al  $ARL_1$  del gráfico  $T^2$ , volviéndose más eficiente cuando los cambios en las medias resultan en distancias de Mahalanobis mayores a 0.1, tal como se puede apreciar en el escenario (b). Cuando el vector de cambios en las medias de las variables es  $d^* = (0.5, 0.5)$  desviaciones estándar, es decir una distancia de Mahalanobis de 0.55, se presenta una reducción en el  $ARL_1$  del gráfico de control LCN del 57.25% en relación al gráfico  $T^2$ .

Igualmente, para  $p=3$  y  $p=4$ , el gráfico LCN presenta mejores rendimientos que la carta de control  $T^2$ , con reducciones en el  $ARL_1$  del más del 40% cuando los cambios en desviaciones estándar del vector de medias resultan en distancias de Mahalanobis mayores o iguales a 0.25.

Por ejemplo, en el escenario (d) cuando el vector de cambios en las medias es  $d^* = (0.5, 0.2, 0.5)$  desviaciones estándar y la distancia de Mahalanobis es de 0.82 (Véase Apéndice D), se produce una reducción del  $ARL_1$  del 63.23%; de manera similar en el escenario (g), cuando el vector de cambios es  $d^* = (0.2, 0.1, 0.1, 0.2)$  y la distancia de Mahalanobis es 0.36 (Véase Apéndice G), se obtiene una mejora del 47%.

Adicionalmente, se compararon los rendimientos para  $p = 3$  y  $p = 4$  entre el gráfico LCPC y  $T^2$ ; considerando para la primera carta mencionada el uso de al menos las dos primeras componentes que expliquen como mínimo el 80% de variabilidad de las variables originales, y para la carta de control  $T^2$ , el uso de las variables originales.

Se puede apreciar que el LCPC control chart (80% de variabilidad) presenta una mejora en el rendimiento en el 97% de los casos analizados. Por ejemplo, en el escenario (h), al querer explicar al menos el 80% de variabilidad, se deben usar dos componentes principales para



el cálculo del estadístico LCPC; además un cambio en las medias de las variables originales de  $d^* = (0.75, 0.5, 0.5, 0.75)$  provoca un cambio en las medias de las dos primeras componentes de  $d^* = (-1.25, -0.04)$  desviaciones estándar y un  $ARL_1$  de 78.92, inferior en 67.14% al obtenido con el gráfico  $T^2$  (Véase Apéndice H).

Por otro lado, los casos en los que no fue mejor el gráfico de control LCPC ocurren cuando las medias originales tienen los siguientes cambios en desviaciones estándar:  $d^* = (1.5, 1.8, 1, 0.2)$  y  $d^* = (2, 1.2, 1.5, 1)$ , es decir, a una distancia de Mahalanobis cercana a 2.5 (Véase Apéndice G); sin embargo, el rendimiento es bastante similar a la carta de control  $T^2$ . En este análisis se han obtenido también los rendimientos para el gráfico LCPC con el 100% de variabilidad de variables originales, el cual tiene un rendimiento similar a la carta de control LCN en la mayoría de los casos.

### 3.3. Análisis de sensibilidad

Como el gráfico de control LCN presentó un buen desempeño, en este apartado se analiza la sensibilidad del mismo. El gráfico de control LCN (Linear combination of Normal Variables) tiene como finalidad obtener los parámetros que detecten rápidamente un cambio dado  $d^* = (d_1, d_2, d_3, \dots, d_p)$  es por ello, que dicho gráfico mostrará un buen desempeño para ese cambio, pero qué pasaría si en base a esos parámetros óptimos existen cambios diferentes a  $d^*$ . Para lo cual, el análisis de sensibilidad nos ayudará a comparar los distintos valores del  $ARL_1$  para diferentes vectores de cambios de dos gráficos de control LCN optimizados para  $d_1^*$  y  $d_2^*$ , respectivamente.

Por la facilidad que resulta el trabajo en dos dimensiones, el análisis mencionado se llevará a cabo para la combinación lineal de dos variables en casos diferentes, es decir, cuándo dichas variables tengan una correlación alta y mediana positiva o negativamente. Se han

obtenido los parámetros de los gráficos de control optimizados para  $d_1^*$  y  $d_2^*$ , y con ellos se han calculado los valores del  $ARL_1$  considerando diversos cambios en las medias de las variables originales que están entre 0.5 y 3 desviaciones estándar.

Al trabajar con dos variables estandarizadas, se entiende que la media y la varianza estarán establecidas con un valor de cero y uno, respectivamente; para lo cual, se analizaron varios casos con distintos valores de correlación. En la ejemplificación del análisis de sensibilidad se mostrarán los casos con una correlación alta de 0.85 y -0.85 además, de una correlación media de 0.35 y -0.45 para el mencionado par de variables.

El gráfico de control LCN se ha optimizado teniendo en cuenta distintos valores para el  $ARL_0$  los casos de correlación descritos en el anterior párrafo y los siguientes vectores de cambios:

**Tabla 6. Vectores de cambios considerados para el análisis de sensibilidad del gráfico de control LCN.**

(a) $d^* = (0.25, 0.25)$	(f) $d^* = (-0.60, -1.20)$
(b) $d^* = (2.00, 2.00)$	(g) $d^* = (-0.25, 0.50)$
(c) $d^* = (-0.25, -0.25)$	(h) $d^* = (-0.75, 1.25)$
(d) $d^* = (-2.00, -2.00)$	(i) $d^* = (-0.90, 0.90)$
(e) $d^* = (-1.20, -2.00)$	(j) $d^* = (-1.50, 1.50)$

*Fuente: Elaboración propia*

Cabe destacar, que dichos vectores de cambios han sido escogidos en función al valor de correlación a analizar. A continuación, se detalla los resultados obtenidos en el análisis de sensibilidad, considerando los criterios mencionados en este apartado.

**Tabla 7. ARLs fuera de control para los gráficos LCN optimizados para los vectores de cambios:  $a = (0.25, 0.25)$ , y  $b = (2.00, 2.00)$ ; con una correlación positiva de 0.85 y  $ARL_0 = 465$ . Los valores entre paréntesis son las mediciones obtenidas para el cambio (b).**

		Cambios en $X_2$					
			0.5	1	1.5	2	2.5
Cambios en $X_1$	0.5	102.026 (103.01)	53.227 (48.446)	29.415 (24.635)	17.203 (13.524)	10.633 (7.998)	6.937 (5.083)
	1	51.81 (58.071)	28.704 (28.962)	16.827 (15.600)	10.426 (9.058)	6.818 (5.656)	4.697 (3.786)
	1.5	28.012 (34.214)	16.462 (18.081)	10.224 (10.305)	6.701 (6.32)	4.627 (4.159)	3.359 (2.928)
	2	16.106 (21.055)	10.027 (11.778)	6.587 (7.094)	4.558 (4.588)	3.316 (3.177)	2.53 (2.345)
	2.5	9.834 (13.524)	6.475 (7.998)	4.491 (5.083)	3.274 (3.461)	2.503 (2.515)	2.00 (1.943)
	3	6.366 (9.058)	4.425 (5.656)	3.233 (3.786)	2.476 (2.708)	1.983 (2.061)	1.654 (1.66)

*Fuente: Elaboración propia*

La Tabla 7 muestra los ARL's fuera de control del gráfico optimizado para un par de variables con una alta correlación positiva y vectores de cambio (a) y (b). Los resultados de ambas cartas de control son similares cuando las variaciones en  $X_1$  y  $X_2$  son iguales, esto lo podemos notar en la diagonal principal de la tabla. Notemos que, si el cambio en  $X_1$  es pequeño y el cambio en  $X_2$  es ligeramente superior o viceversa, los valores del ARL's van a diferir un poco para cada gráfico.

**Tabla 8. ARLs fuera de control para los gráficos LCN optimizados para los vectores de cambios:  $c = (-0.25, -0.25)$ , y  $d = (-2.00, -2.00)$ ; con una correlación positiva de 0.85,  $ARL_0 = 465$ . Los valores entre paréntesis son las mediciones obtenidas para el cambio (d).**

		Cambios en $X_2$					
		-0,5	-1	-1,5	-2	-2,5	-3
Cambios en $X_1$	-0,5	102.96 (103.188)	53.23 (42.914)	29.195 (19.868)	16.97 (10.21)	10.441 (5.804)	6.79 (3.633)
	-1	52.684 (66.478)	28.923 (29.135)	16.828 (14.196)	10.363 (7.66)	6.745 (4.569)	4.632 (2.991)
	-1,5	28.654 (43.938)	16.687 (20.281)	10.286 (10.392)	6.701 (5.89)	4.606 (3.677)	3.332 (2.513)
	-2	16.548 (29.785)	10.21 (14.47)	6.657 (7.791)	4.579 (4.631)	3.316 (3.023)	2.521 (2.154)
	-2,5	10.134 (20.703)	6.614 (10.578)	4.553 (5.979)	3.3 (3.722)	2.51 (2.537)	2.00 (1.881)
	-3	6.571 (14.75)	4.527 (7.919)	3.284 (4.694)	2.501 (3.056)	1.99 (2.172)	1.657 (1.672)

*Fuente: Elaboración propia*

Algo similar ocurre para los resultados de los gráficos de control optimizados para los vectores de cambios (c) y (d), ilustrados en la Tabla 8; aunque existe una ligera deficiencia cuando dichos cambios se alejan el uno con el otro o viceversa. Por ejemplo, en el caso de que  $d_1 = -1.5$  y  $d_2 = -0.5$  el valor de  $ARL_1$  para el primer gráfico es de 29.195 que es mayor en un 46.94% que el  $ARL_1$  obtenido para el mismo cambio, en el otro gráfico.

**Tabla 9. ARLs fuera de control para los gráficos LCN optimizados para los vectores de cambios:  $e = (-1.20, -2.00)$ , y  $f = (-0.60, -1.20)$ ; con una correlación positiva de 0.35,  $ARL_0 = 750$ . Los valores entre paréntesis son las mediciones obtenidas para el cambio (f).**

		Cambios en $X_2$					
		-0,5	-1	-1,5	-2	-2,5	-3
Cambios en $X_1$	-0,5	126.848 (77.235)	45.126 (29.383)	18.541 (12.885)	8.758 (6.481)	4.729 (3.713)	2.896 (2.403)
	-1	77.235 (88.096)	29.383 (31.139)	12.885 (12.919)	6.481 (6.255)	3.713 (3.505)	2.403 (2.249)
	-1,5	48.518 (60.168)	19.723 (22.503)	9.222 (9.86)	4.931 (5.029)	2.992 (2.958)	2.041 (1.984)
	-2	31.435 (41.918)	13.641 (16.579)	6.792 (7.666)	3.855 (4.114)	2.473 (2.537)	1.773 (1.775)
	-2,5	20.997 (29.785)	9.716 (12.449)	5.145 (6.069)	3.094 (3.423)	2.093 (2.21)	1.572 (1.61)
	-3	14.452 (21.581)	7.123 (9.525)	4.005 (4.891)	2.546 (2.896)	1.811 (1.953)	1.422 (1.478)

*Fuente: Elaboración propia*

Para el caso del par de variables con una correlación positiva de 0.35 y vectores de cambio (e) y (f), los resultados obtenidos en la Tabla 9, son totalmente distintos. Como las magnitudes de los vectores de cambio considerados en la optimización son moderadas, cuando los valores de  $d_1$  y  $d_2$  son grandes, los gráficos de control optimizados muestran rendimientos muy similares pero en cambio; cuando los cambios inesperados en las variables  $X_1$  y  $X_2$  se alejan de los vectores de cambio (e) y (f), los rendimientos de ambos gráficos son distintos; por ejemplo para el caso de que  $d_1 = -0.5$  y  $d_2 = -0.5$  el valor  $ARL_1$  del primer gráfico es mayor en un 64.23% en comparación con el rendimiento obtenido en el segundo gráfico optimizado para el cambio (f).

**Tabla 10. ARLs fuera de control para los gráficos LCN optimizados para los vectores de cambios:  $g = (-0.25, 0.50)$ , y  $h = (-0.75, 1.25)$ ; con una correlación negativa de  $-0.45$ ,  $ARL_0 = 900$ . Los valores entre paréntesis son las mediciones obtenidas para el cambio (h).**

		Cambios en $X_2$					
		0.5	1	1.5	2	2.5	3
Cambios en $X_1$	-0,5	179.136 (171.464)	50.594 (50.94)	17.518 (18.275)	7.381 (7.865)	3.745 (4.024)	2.257 (2.418)
	-1	158.324 (138.21)	45.56 (42.387)	16.065 (15.683)	6.888 (6.952)	3.552 (3.657)	2.172 (2.254)
	-1,5	140.172 (112.003)	41.10 (35.455)	14.757 (13.526)	6.437 (6.174)	3.374 (3.338)	2.093 (2.109)
	-2	124.314 (91.121)	37.138 (29.811)	13.577 (11.725)	6.025 (5.509)	3.209 (3.06)	2.02 (1.981)
	-2,5	110.44 (74.739)	33.614 (25.195)	12.512 (10.214)	5.648 (4.939)	3.057 (2.818)	1.952 (1.868)
	-3	98.282 (61.54)	30.474 (21.403)	11.549 (8.941)	5.303 (4.448)	2.915 (2.605)	1.888 (1.768)

Fuente: Elaboración propia

En la Tabla 10 se muestran los rendimientos para los gráficos de control optimizados para un par de variables con correlación negativa  $-0.45$  y para los vectores de cambios (g) y (h), que consideran una variación pequeña en la media de la primera variable. Se puede observar que cuando  $d_1$  es cercano a cero y  $d_2$  varía positivamente, los rendimientos se muestran similares para los gráficos de control optimizados a dichos vectores de cambios; por ejemplo, cuando  $d_1 = -0.5$  y  $d_2 = 2.5$  el  $ARL_1$  muestra un valor cercano a 4 para los dos gráficos, algo que no se muestra al momento de que  $d_2$  es cercano a cero y constante a decrementos de  $d_1$ , ocurriendo totalmente lo contrario a lo anteriormente mencionado.

**Tabla 11. ARLs fuera de control para los gráficos LCN optimizados para los vectores de cambios:  $i = (-0.90, 0.90)$ , y  $j = (-1.50, 1.50)$ ; con una correlación negativa de  $-0.85$ ,  $ARL_0 = 500$ . Los valores entre paréntesis son las mediciones obtenidas para el cambio (j).**

		Cambios en $X_2$					
		0,5	1	1,5	2	2,5	3
Cambios en $X_1$	-0,5	110.95 (112.346)	78.98 (94.434)	57.058 (79.681)	41.83 (67.488)	31.127 (57.377)	23.497 (48.964)
	-1	41.84 (36.971)	31.127 (31.886)	23.497 (27.601)	17.99 (23.981)	13.978 (20.912)	11.011 (18.302)
	-1,5	17.995 (14.469)	13.98 (12.793)	11.01 (11.351)	8.795 (10.108)	7.121 (9.032)	5.844 (8.099)
	-2	8.795 (6.691)	7.121 (6.058)	5.844 (5.504)	4.895 (5.017)	4.092 (4.588)	3.491 (4.209)
	-2,5	4.859 (3.625)	4.092 (3.356)	3.491 (3.117)	3.014 (2.903)	2.634 (2.713)	2.328 (2.542)
	-3	3.014 (2.274)	2.634 (2.148)	2.328 (2.148)	2.081 (2.035)	1.881 (1.933)	1.717 (1.758)

*Fuente: Elaboración propia*

Para el caso del par de variables con una correlación negativa alta y vectores de cambio (i) y (j) (Véase Tabla 11), los valores del ARL obtenidos muestran comportamientos similares a la Tabla 7, llegando a la misma conclusión.

En general, para cualquier cambio establecido para los diferentes LCN gráficos de control optimizados para un par de variables con una alta correlación; se obtiene rendimientos similares, sin existir mucha variación entre uno y otro; por lo tanto, notamos un grado de insensibilidad para estos casos, siempre y cuando el cambio real del proceso tenga la misma orientación (incremento o decremento) que el vector de cambios considerado en la optimización.

Pero en cambio, cuando las correlaciones son medianas, ya sea positiva o negativa, los valores de  $ARL_s$  suelen variar y diferenciarse con otros gráficos LCN a medida que los cambios inesperados se alejan del vector de cambio por el cual el gráfico de control fue optimizado, mostrando así un grado de sensibilidad.

Para estos casos, se recomienda al usuario establecer un vector de cambio general que siempre suele darse en el proceso, esperando que los cambios no previstos sean cercanos a dicho vector.



# CAPÍTULO 4

## 4. DISCUSIÓN Y CONCLUSIONES

### 4.1. Conclusiones

- Se logró desarrollar dos estadísticos: uno basado en la combinación lineal de variables normales y otro fundamentado en la combinación lineal de al menos las dos primeras componentes principales, que expliquen como mínimo el 80% de la variabilidad total; para el diseño de los gráficos de control LCN y LCPC, respectivamente.
- Se logró realizar la programación de algoritmos genéticos en Microsoft Visual Basic Community©, que permitieron conocer los parámetros óptimos para los gráficos propuestos en esta investigación.
- Para el control de calidad en un proceso de producción de galletas, se pudo evidenciar la aplicabilidad y mejora en el rendimiento de los gráficos LCN y LCPC, en comparación con el  $T^2$  de Hotelling.
- El gráfico de control LCN presentó mejores rendimientos que el esquema multivariado  $T^2$  de Hotelling, considerando  $p=2$ ,  $p=3$  y  $p=4$  variables con diversos cambios en las medias, distintas correlaciones y valores diferentes para el  $ARL_0$ .
- Se obtuvo mejores rendimientos del gráfico LCPC (80% variabilidad) en comparación con el  $T^2$ , en un 97% de los casos analizados.
- El gráfico LCN mostró un alto grado de sensibilidad cuando las correlaciones (positiva o negativa) del conjunto de variables son medianas ya que, los valores de  $ARL_1$  suelen variar y diferenciarse con otros gráficos LCN a medida que los cambios inesperados se alejan del vector de cambio por el cual el gráfico de control fue optimizado.

- Se observó un grado de insensibilidad en el gráfico de control LCN cuando las correlaciones del conjunto de variables son altas ya que, al realizar diferentes comparaciones, se presentaron rendimientos similares entre un gráfico y otro, sin existir mucha variación.

#### **4.2. Recomendaciones**

- Trabajar con el conjunto de variables originales y con el gráfico de control LCN por facilidad en la interpretación, ya que este esquema presenta en su totalidad rendimientos superiores al gráfico LCPC, a pesar de que este considera al menos el 80% de la variabilidad total.
- Al utilizar el software diseñado para el gráfico de control LCN y LCPC se debe establecer, un vector de cambio general que siempre suele darse en el proceso, esperando que los cambios no previstos sean cercanos a dicho vector.

# BIBLIOGRAFÍA

- Aparisi, F., & de Luna, M. A. (2009). Synthetic -  $\bar{X}$  control charts optimized for in-control and out-of-control regions. *Computers & Operations Research*, 36(12), 3204-3214.
- Aparisi, F., & García-Díaz, J. (2001). Aumento de la potencia del gráfico de control multivariante  $T^2$  de Hotelling utilizando señales adicionales de falta de control. *Revista Estadística Española.*, 43(148), 171-188.
- Aparisi, F., & García-Díaz, J. (2007). Design and Optimization of EWMA Control Charts for In-control, Indifferent, and Out-of-control Regions. *Computers & Operations Research (Revista JCR).*, 34(7), 2096-2108.
- Aparisi, F., de Luna, M., & Epprecht, E. (2010). Optimisation of a set of or principal components control charts using genetic algorithms. *International Journal of Production Research*, 48(18), 5345-5361.
- Barahona, J. (2016). PROYECTO DE GRADUACIÓN: Desarrollo de un paquete en R para Gráficos de control multivariantes. Guayaquil: ESPOL.
- Cook, D. F., & Chiu, C. C. (1998). Using radial basis function neural network to recognize shift in manufacturing process. *IIE Transactions*, 30(3), 227-234.
- Crosier, R. B. (1988). Multivariate generalizations of cumulative sum quality-control schemes. *Technometrics*, 30(3), 291-303.
- Darwin, C. (2003). *El origen de las especies* (Vol. No. 575). (J. d. Rojas, Ed.)
- Epprecht, E. K., García-Bustos, S., & Aparisi, F. (2013). Optimal linear combination of Poisson variables for statistical process control. *Computers & Operations Research*, 40, 3021-3032.

- García-Bustos, S. L. (2014). TESIS DOCTORAL: Control Multivariante Estadístico de Variables Discretas tipo Poisson. Valencia: Universidad politécnica de Valencia.
- García-Bustos, S., Aparisi, F., & Epprecht, E. K. (2014). Optimal EWMA of Linear Combination of Poisson Variables for Multivariate Statistical Process Control. *International Journal of Production Research*, 53(14), 4141-4159.
- García-Bustos, S., Mite, M., & Vera, F. (2015). Control Charts with Variable Dimension for Linear Combination of Poisson Variables. *Quality and Reliability Engineering International*, 32(5), 1741-1755.
- Hayter, A. J., & KNOW-LEUNG, T. S. (1994). Identification and quantification in multivariate quality control problems. *Journal of Quality Technology*, 26(3), 197-208.
- Holland, J. (1992). Algoritmos genéticos. *Investigación y Ciencia*(192), 38-45.
- Hotelling, H. (1947). Multivariate quality control, illustrated by the air testing of simple bombsights. *Techniques of statistical analysis*, 111-184.
- Lowry, C., & Montgomery, D. (1995). A review of multivariate control charts. *IIE transactions*, 27(6), 800-810.
- Lowry, C., Woodall, W., Champ, C., & Rigdon, S. (1992). A multivariate exponentially weighted moving average control chart. *Technometrics*, 34(1), 46-53.
- MacGregor, J., & Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, 3(3), 403-414.
- Montgomery, D. C. (2007). *Introduction to statistical quality control*. (6th ed., Vol. 7). New York: John Wiley & Sons.
- Moujahid, A., Inza, I., & Larrañaga, P. (s.f.). *Tema 2. Algoritmos Genéticos*. Universidad del País Vasco–Euskal Herriko Unibertsitatea, Departamento

de Ciencias de la Computación e Inteligencia Artificial. Obtenido de <http://www.sc.ehu.es/ccwbayes/docencia/mmcc/docs/t2geneticos.pdf>

Murphy, B. (1987). Selecting Out of Control Variables With the  $T^2$  Multivariate Quality Control Procedure. *The Statistician*, 571-581.

Pignatiello, J., & Runger, G. (1990). Comparisons of multivariate CUSUM charts. *Journal of quality technology*, 22(3), 173-186.

R Core, T. (s.f.). R: A language and environment for statistical computing. 2017. Viena, Austria: R Foundation for Statistical Computing. Obtenido de <https://www.R-project.org/>

Rencher, A. (2003). *Methods of Multivariate Analysis* (2da ed., Vol. 492). John Wiley & Sons.

Rojas, M., Hernández, A., & Ruiz, J. (2014). Análisis comparativo del control estadístico de procesos (CEP) univariable y multivariada en un proceso de producción de galletas. *Revista Electrónica de Divulgación de la Investigación*, 8, 41-60.

Runger, G. C., & Montgomery, D. C. (1997). Multivariate and univariate process control: geometry and shift directions. *Quality and Reliability Engineering International*, 13(3), 153-158.

Seber, G. A. (2009). *Multivariate observations* (Vol. 252). John Wiley & Sons.

Shabbak, A., & Midi, H. (2012). An Improvement of the Hotelling Statistic in Monitoring Multivariate Quality Characteristics. . *Mathematical Problems in Engineering*.

Tran, K., Castagliola, P., & Celano, G. (2016). Monitoring the ratio of two normal variables using run rules type control charts. *International Journal of Production Research*, 54(6), 1670-1688.

Wackerly, D., Mendenhall, W., & Scheaffer, R. (2010). *Estadística Matemática con Aplicaciones*. (7th ed.). Mexico D.F.: Cengage Learning.

Wikipedia. (s.f.). Algoritmo Genético. Recuperado el Agosto de 2017, de [https://es.wikipedia.org/wiki/Algoritmo\\_gen%C3%A9tico](https://es.wikipedia.org/wiki/Algoritmo_gen%C3%A9tico)

Zamarrón, A. M., Prado, E., & Luis, F. (2012). Monitoreo y control de un proceso normal multivariado. *Conciencia Tecnológica*, 43, 29-35.

# APÉNDICES

## APÉNDICE A

Número de muestra	$Z_{P.Crudo}$	$Z_{P.Horneado}$	$Z_{Volumen}$	LCN
1	1.685	1.069	1.545	-1.888
2	0.790	0.889	0.710	-1.051
3	0.188	0.379	0.005	-0.217
4	0.545	0.837	0.994	-1.158
5	-1.279	-1.320	-1.231	1.701
6	1.458	0.718	1.366	-1.560
7	1.916	2.284	1.582	-2.513
8	0.552	1.353	2.274	-2.237
9	0.979	1.053	0.577	-1.080
10	-1.274	-1.234	0.360	0.610
11	1.792	1.167	0.985	-1.588
12	1.544	2.015	2.878	-3.161
13	3.162	3.080	2.063	-3.475
14	1.653	0.818	0.620	-1.157
15	0.943	1.308	2.081	-2.179
16	1.308	1.838	1.995	-2.445
17	-0.445	-0.651	0.299	0.198
18	-0.837	-0.521	-0.671	0.870
19	0.272	0.509	0.411	-0.563
20	2.829	3.542	3.186	-4.347
21	0.632	1.263	1.023	-1.388
22	1.059	1.180	0.853	-1.338
23	0.434	0.403	0.781	-0.797
24	0.752	1.377	1.622	-1.863
25	1.234	1.373	1.136	-1.652

Datos simulados para un proceso de horneado fuera de control con cambios en el vector de medias  $d = (1, 1, 1)$ , utilizados para el diseño del gráfico

$$LCN = -0.23 Z_{P.Crudo} - 0.45 Z_{P.Horneado} - 0.66 Z_{Volumen}$$

## APÉNDICE B

Número de muestra	C <sub>1</sub>	C <sub>2</sub>	LCPC
1	-2.468	0.067	<b>0.995</b>
2	-1.382	-0.056	<b>0.547</b>
3	-0.339	-0.177	<b>0.116</b>
4	-1.365	0.321	<b>0.581</b>
5	2.211	0.002	<b>-0.884</b>
6	-2.028	0.122	<b>0.824</b>
7	-3.352	-0.272	<b>1.311</b>
8	-2.383	1.241	<b>1.090</b>
9	-1.517	-0.302	<b>0.574</b>
10	1.281	1.255	<b>-0.374</b>
11	-2.281	-0.461	<b>0.862</b>
12	-3.687	1.029	<b>1.588</b>
13	-4.817	-0.753	<b>1.844</b>
14	-1.791	-0.606	<b>0.650</b>
15	-2.476	0.872	<b>1.086</b>
16	-2.959	0.492	<b>1.238</b>
17	0.484	0.620	<b>-0.126</b>
18	1.167	0.042	<b>-0.462</b>
19	-0.689	0.075	<b>0.284</b>
20	-5.518	0.246	<b>2.234</b>
21	-1.687	0.212	<b>0.698</b>
22	-1.791	-0.151	<b>0.700</b>
23	-0.923	0.299	<b>0.402</b>
24	-2.154	0.600	<b>0.928</b>
25	-2.164	-0.062	<b>0.859</b>

**Apéndice B. Valores de las dos primeras componentes principales  $C_1 = -0.58 Z_{P.Crudo} - 0.60 Z_{P.Horneado} - 0.55 Z_{Volumen}$  y  $C_2 = -0.57 Z_{P.Crudo} - 0.195 Z_{P.Horneado} + 0.80 Z_{Volumen}$  resultantes de los datos simulados, utilizados para el diseño del gráfico  $LCPC = -0.4 C_1 + 0.11 C_2$ .**



## APÉNDICE C

Cambios en las medias		Dist. de Mahalanobis A	Dist. de Mahalanobis B	Dist. de Mahalanobis C	Escenario (a)			Escenario (b)			Escenario (c)		
					T <sup>2</sup>	LCN	% de mejora	T <sup>2</sup>	LCN	% de mejora	T <sup>2</sup>	LCN	% de mejora
X1	X2				ARL(d=d*)	ARL(d=d*)		ARL(d=d*)	ARL(d=d*)		ARL(d=d*)	ARL(d=d*)	
0.00	0.00	0.00	0.00	0.00	300	300		500	500		400	400	
0.10	0.00	0.17	0.13	0.10	277.73	183.85	-33.80%	474.39	331.37	-30.15%	387.25	289.53	-25.23%
0.10	0.10	0.11	0.11	0.12	290.77	219.47	-24.52%	481.88	354.67	-26.40%	382.32	272.91	-28.62%
0.00	0.20	0.33	0.23	0.21	225	116.94	-48.03%	426.08	222.41	-47.80%	352.63	212.05	-39.87%
0.20	0.20	0.21	0.22	0.25	256.66	161.91	-36.92%	433.21	254.26	-41.31%	336.2	188.01	-44.08%
0.50	0.00	0.83	0.66	0.52	82.35	36.22	-56.02%	188.09	76.32	-59.42%	206.4	90.03	-56.38%
0.50	0.50	0.53	0.55	0.62	158.26	70.02	-55.76%	238.58	101.99	-57.25%	168.1	69.82	-58.47%
0.75	0.00	1.25	0.99	0.79	33.63	17.85	-46.92%	87.55	43.9	-49.86%	115.95	46.74	-59.69%
0.75	0.75	0.79	0.83	0.93	90.46	36.85	-59.26%	127.36	49.94	-60.79%	83.64	33.25	-60.25%
1.00	0.00	1.67	1.32	1.05	14.92	10.38	-30.43%	41.8	18.64	-55.41%	64.13	25.66	-59.99%
1.00	1.00	1.05	1.10	1.24	50.91	20.75	-59.24%	67.42	26.62	-60.52%	42.07	17.39	-58.66%
1.50	0.00	2.50	1.97	1.57	4.15	2.41	-41.93%	11.51	5.56	-51.69%	21.25	9.23	-56.56%
1.50	1.50	1.58	1.65	1.86	17.48	7.77	-55.55%	20.9	9.2	-55.98%	12.43	5.86	-52.86%
2.00	2.00	2.11	2.20	2.48	7.13	3.68	-48.39%	7.89	3.68	-53.36%	4.77	2.91	-38.99%

**Comparación de rendimientos utilizando dos variables.**

## APÉNDICE D

Cambios en las medias			Cambios en las medias de las componentes			Dist. de Mahalanobis	Gráfico T <sup>2</sup>	Gráfico LCPC (100%)	Gráfico LCPC (80%)	Gráfico LCN	% de mejora gráfico LCN	% de mejora gráfico LCPC (80%)
X1	X2	X3	C1	C2	C3		ARL(d=d*)	ARL(d=d*)	ARL(d=d*)	ARL(d=d*)		
0	0	0	0.00	0.00	0.00	0.00	400	400	400	400		
0.1	0.05	0	-0.09	-0.06	0.03	0.13	383.6	265.01	294.32	273.32	-28.75%	-23.27%
0.1	0.1	0.1	-0.17	0.01	0.00	0.12	385.5	281.19	281.58	280.01	-27.36%	-26.96%
0	0	0.2	-0.10	0.17	0.03	0.25	347.43	189.63	196.17	191.84	-44.78%	-43.54%
0.2	0.1	0.2	-0.28	0.05	0.08	0.30	328.55	165.87	221.58	194.81	-40.71%	-32.56%
0.5	0.6	0.3	-0.82	-0.13	-0.08	0.61	203.01	72.57	79.71	73.9	-63.60%	-60.74%
0.5	0.2	0.5	-0.68	0.14	0.24	0.82	132.77	47.4	100.68	48.82	-63.23%	-24.17%
0.75	0.4	0.1	-0.75	-0.36	0.21	0.89	115.11	36.67	57.7	37.38	-67.53%	-49.87%
0.75	0.5	0.75	-1.14	0.16	0.22	0.98	94.81	31.29	47.04	36.43	-61.58%	-50.38%
1	1	0.6	-1.52	-0.21	-0.01	1.04	83.88	26.03	26.08	26.06	-68.93%	-68.91%
1	1	1	-1.73	0.13	0.04	1.16	65.64	20.3	20.7	20.75	-68.39%	-68.46%
1.5	1.5	1.5	-2.59	0.20	0.06	1.74	20.67	7.13	7.24	7.64	-63.04%	-64.97%
2	1.2	1	-2.44	-0.38	0.55	2.25	8.56	5.26	7.61	5.3	-38.08%	-11.10%
2	1.5	2	-3.14	0.39	0.45	2.48	6.11	3.46	4.08	3.52	-42.39%	-33.22%

Comparación de rendimientos utilizando tres variables. Escenario (d), los coeficientes de las componentes principales son  $\alpha_{1a} = -0.60$ ,  $\alpha_{2a} = -0.62$ ,  $\alpha_{3a} = -0.51$ ;  $\alpha_{1b} = -0.47$ ,  $\alpha_{2b} = -0.25$ ,  $\alpha_{3b} = 0.85$ ;  $\alpha_{1c} = 0.65$ ,  $\alpha_{2c} = -0.74$ ,  $\alpha_{3c} = 0.13$ , y sus varianzas son  $\lambda_a = 2.31$ ,  $\lambda_b = 0.55$ ,  $\lambda_c = 0.14$

## APÉNDICE E

Cambios en las medias			Cambios en las medias de las componentes			Dist. de Mahalanobis	Gráfico T <sup>2</sup>	Gráfico LCPC (100%)	Gráfico LCPC (80%)	Gráfico LCN	% de mejora gráfico LCN	% de mejora gráfico LCPC (80%)
X1	X2	X3	C1	C2	C3		ARL(d=d*)	ARL(d=d*)	ARL(d=d*)	ARL(d=d*)		
0	0	0	0.00	0.00	0.00	0.00	600	600	600	600		
0.1	0.05	0	-0.07	0.08	-0.02	0.11	576.08	421.48	431.76	427.01	-25.88%	-25.05%
0.1	0.1	0.1	-0.17	0.03	0.01	0.13	570.08	398.08	401.87	396.12	-30.52%	-29.51%
0	0	0.2	-0.12	-0.08	0.13	0.31	482.19	237.5	397.75	279.11	-42.12%	-17.51%
0.2	0.1	0.2	-0.27	0.08	0.10	0.30	464.57	249.48	306.73	265.68	-42.81%	-33.98%
0.5	0.6	0.3	-0.79	0.23	-0.15	0.69	244.48	81.18	97.49	87.27	-64.30%	-60.12%
0.5	0.2	0.5	-0.65	0.21	0.28	0.78	204.11	65.21	132.57	82.41	-59.62%	-35.05%
0.75	0.4	0.1	-0.63	0.57	-0.08	0.79	200.23	63.13	65.92	65.36	-67.36%	-67.08%
0.75	0.5	0.75	-1.11	0.28	0.27	1.02	119.94	36.04	54.49	43.48	-63.75%	-54.57%
1	1	0.6	-1.45	0.49	-0.15	1.21	79.51	24.12	27.11	25.44	-68.00%	-65.90%
1	1	1	-1.70	0.32	0.12	1.29	66.33	20.46	20.84	20.6	-68.94%	-68.58%
1.5	1.5	1.5	-2.55	0.49	0.17	1.94	17.95	6.66	6.78	6.51	-63.73%	-62.23%
2	1.2	1	-2.24	1.19	0.16	2.09	13.7	5.14	6.18	5.08	-62.92%	-54.89%
2	1.5	2	-3.06	0.72	0.60	2.66	5.58	2.96	4.05	2.81	-49.64%	-27.42%

**Comparación de rendimientos utilizando tres variables. Escenario (e), los coeficientes de las componentes principales son  $\alpha_{1a} = -0.41$ ,  $\alpha_{2a} = -0.67$ ,  $\alpha_{3a} = -0.62$ ;  $\alpha_{1b} = 0.89$ ,  $\alpha_{2b} = -0.15$ ,  $\alpha_{3b} = -0.42$ ;  $\alpha_{1c} = 0.19$ ,  $\alpha_{2c} = -0.73$ ,  $\alpha_{3c} = 0.66$ , y sus varianzas son  $\lambda_a = 1.93$ ,  $\lambda_b = 0.84$ ,  $\lambda_c = 0.23$**

## APÉNDICE F

Cambios en las medias			Cambios en las medias de las componentes			Dist. de Mahalanobis	Gráfico T <sup>2</sup>	Gráfico LCPC (100%)	Gráfico LCPC (80%)	Gráfico LCN	% de mejora gráfico LCN	% de mejora gráfico LCPC (80%)
X1	X2	X3	C1	C2	C3		ARL(d=d*)	ARL(d=d*)	ARL(d=d*)	ARL (d=d*)		
0	0	0	0.00	0.00	0.00	0.00	800	800	800	800		
0.1	0.05	0	-0.09	-0.02	0.07	0.12	760.81	545.87	631.86	543.56	-28.56%	-16.95%
0.1	0.1	0.1	-0.17	-0.03	0.00	0.14	750.57	510.15	510.22	527.61	-29.71%	-32.02%
0	0	0.2	-0.13	0.07	-0.14	0.22	689.09	386.72	530.42	349.03	-49.35%	-23.03%
0.2	0.1	0.2	-0.30	0.03	0.00	0.24	687.59	372.33	373.85	376.52	-45.24%	-45.63%
0.5	0.6	0.3	-0.77	-0.31	0.11	0.71	304.53	97.24	100.97	90	-70.45%	-66.84%
0.5	0.2	0.5	-0.73	0.13	0.00	0.59	390.06	131.9	133.26	134.37	-65.55%	-65.84%
0.75	0.4	0.1	-0.72	-0.14	0.44	0.84	229.25	73.59	135.7	56.05	-75.55%	-40.81%
0.75	0.5	0.75	-1.18	0.01	-0.02	0.93	186.55	56.5	54.74	62.13	-66.70%	-70.66%
1	1	0.6	-1.46	-0.43	0.23	1.28	83.92	25.84	26.61	23.41	-72.10%	-68.29%
1	1	1	-1.71	-0.29	-0.05	1.39	66.18	19.65	19.46	25.89	-60.88%	-70.60%
1.5	1.8	0	-1.76	-1.25	0.94	2.31	10.98	6.23	7.61	5.05	-54.01%	-30.69%
1.5	1.5	1.5	-2.57	-0.44	-0.07	2.09	16.19	5.74	5.76	7.38	-54.42%	-64.42%
2	1.5	2	-3.21	-0.13	-0.06	2.54	7.54	3.31	3.24	4.14	-45.09%	-57.03%

Comparación de rendimientos utilizando tres variables. Escenario (f), los coeficientes de las componentes principales son  $\alpha_{1a} = -0.65$ ,  $\alpha_{2a} = -0.43$ ,  $\alpha_{3a} = -0.63$ ;  $\alpha_{1b} = 0.25$ ,  $\alpha_{2b} = -0.9$ ,  $\alpha_{3b} = 0.36$ ;  $\alpha_{1c} = 0.72$ ,  $\alpha_{2c} = -0.08$ ,  $\alpha_{3c} = -0.69$ , y sus varianzas son  $\lambda_a = 1.59$ ,  $\lambda_b = 0.87$ ,  $\lambda_c = 0.55$

## APÉNDICE G

Cambios en las medias				Dist. de Mahalanobis	Gráfico T <sup>2</sup>	Gráfico LCPC (100%)	Gráfico LCPC (80%)	Gráfico LCN	% de mejora gráfico LCN	% de mejora gráfico LCPC (80%)
X1	X2	X3	X4		ARL (d=d*)	ARL (d=d*)	ARL (d=d*)	ARL (d=d*)		
0	0	0	0	0.00	600	600	600	600		
0.1	0.1	0.1	0.1	0.11	584.94	417.29	419.73	404.17	-30.90%	-28.24%
0.2	0.1	0.1	0.2	0.36	473.7	207.68	333.6	252.99	-46.59%	-29.58%
0.5	0.6	0.3	0.25	0.72	266.59	82.44	143.38	89.89	-66.28%	-46.22%
0.75	0.4	0.6	0.15	0.99	155.67	50.94	74.69	51.11	-67.17%	-52.02%
0.75	0.5	0.5	0.75	1.08	127.53	35.21	71.83	53.28	-58.22%	-43.68%
1	1	1	0.5	1.09	125.36	31.01	32.78	27.2	-78.30%	-73.85%
1	1	0.5	1	1.33	75.64	20.12	35.08	13.37	-82.32%	-53.62%
1.5	1.8	1	0.2	2.46	9.25	3.33	9.93	6.68	-27.78%	7.35%
1.5	1.5	0.5	1.5	2.29	12.09	4.37	10.55	9.27	-23.33%	-12.74%
2	1.2	1.5	1	2.47	9.03	5.63	9.51	6.24	-30.90%	5.32%
2	1.5	1.8	2	2.59	7.5	3.5	5.12	4.52	-39.73%	-31.73%

Comparación de rendimientos utilizando cuatro variables. Escenario (g), los coeficientes de las componentes principales son  $\alpha_{1a} = -0.53, \alpha_{2a} = -0.54, \alpha_{3a} = -0.50, \alpha_{4a} = -0.44; \alpha_{1b} = -0.29, \alpha_{2b} = -0.03, \alpha_{3b} = -0.40, \alpha_{4b} = 0.87; \alpha_{1c} = -0.35, \alpha_{2c} = -0.53, \alpha_{3c} = 0.74, \alpha_{4c} = 0.21; \alpha_{1d} = 0.72, \alpha_{2d} = -0.66, \alpha_{3d} = -0.17, \alpha_{4d} = 0.14$ , y sus varianzas son  $\lambda_a = 3.22, \lambda_b = 0.51, \lambda_c = 0.19, \lambda_d = 0.08$

## APÉNDICE H

Cambios en las medias				Dist. de Mahalanobis	Gráfico T <sup>2</sup>	Gráfico LCPC (100%)	Gráfico LCPC (80%)	Gráfico LCN	% de mejora gráfico LCN	% de mejora gráfico LCPC (80%)
X1	X2	X3	X4		ARL (d=d*)	ARL (d=d*)	ARL (d=d*)	ARL (d=d*)		
0	0	0	0	0.00	800	800	800	800		
0.1	0.05	0	0	0.12	776.97	543.24	598.75	547.15	-29.58%	-22.94%
0.1	0.1	0.1	0.1	0.13	774.5	531.7	532.29	523.27	-32.44%	-31.27%
0	0	0.2	0.2	0.26	701.66	353.41	362.57	359.02	-48.83%	-48.33%
0.2	0.1	0.1	0.2	0.26	703.03	359.32	433.66	349.77	-50.25%	-38.32%
0.5	0.6	0.3	0	0.71	350.35	102.63	120.65	109.98	-68.61%	-65.56%
0.5	0.2	0.1	0.5	0.75	319.92	94.75	222.29	98.98	-69.06%	-30.52%
0.75	0.4	0.6	0	1.03	180.09	45.8	122.42	47.92	-73.39%	-32.02%
0.75	0.5	0.5	0.75	0.89	240.16	64.58	78.92	60.34	-74.88%	-67.14%
1	1	1	0.5	1.27	108.1	26.53	33.88	28.08	-74.02%	-68.66%
1	1	0.5	1	1.31	98.47	26.09	34.9	24.03	-75.60%	-64.56%
1.5	1.8	1	0.75	1.85	31.85	9.01	9.34	9.06	-71.55%	-70.68%
1.5	1.5	0.5	1.5	2.10	19.72	6.36	12.46	6.05	-69.32%	-36.82%
2	1.5	1.8	2	2.37	12.39	4.23	4.41	4.05	-67.31%	-64.41%

**Apéndice H. Comparación de rendimientos utilizando cuatro variables. Escenario (h), los coeficientes de las componentes principales son  $\alpha_{1a} = -0.53, \alpha_{2a} = -0.48, \alpha_{3a} = -0.50, \alpha_{4a} = -0.49; \alpha_{1b} = 0.39, \alpha_{2b} = 0.59, \alpha_{3b} = -0.46, \alpha_{4b} = -0.53; \alpha_{1c} = -0.14, \alpha_{2c} = 0.18, \alpha_{3c} = -0.68, \alpha_{4c} = 0.69; \alpha_{1d} = 0.74, \alpha_{2d} = -0.62, \alpha_{3d} = -0.25, \alpha_{4d} = 0.07$ , y sus varianzas son  $\lambda_a = 2.52, \lambda_b = 0.85, \lambda_c = 0.34, \lambda_d = 0.28$ .**

## APÉNDICE I

Cambios en las medias				Dist. de Mahalanobis	Gráfico T <sup>2</sup>	Gráfico LCPC (100%)	Gráfico LCPC (80%)	Gráfico LCN	% de mejora gráfico LCN	% de mejora gráfico LCPC (80%)
X1	X2	X3	X4		ARL (d=d*)	ARL (d=d*)	ARL (d=d*)	ARL (d=d*)		
0.00	0.00	0.00	0.00	0.00	1000	1000	1000	1000		
0.10	0.05	0.00	0.00	0.13	966.72	657.56	721.57	678.92	-29.77%	-25.36%
0.10	0.10	0.10	0.10	0.15	956.69	622.7	623.41	624.73	-34.70%	-34.84%
0.00	0.00	0.20	0.20	0.32	817.56	364.48	562.34	367.93	-55.00%	-31.22%
0.20	0.10	0.10	0.20	0.24	891.42	462.01	438.57	420.93	-52.78%	-50.80%
0.50	0.60	0.30	0.10	0.82	342.78	89.78	189.59	106.43	-68.95%	-44.69%
0.50	0.20	0.10	0.50	0.60	528.75	157.83	163.11	126.31	-76.11%	-69.15%
0.75	0.40	0.60	0.00	1.02	219.97	57.72	72.28	75.39	-65.73%	-67.14%
0.75	0.50	0.50	0.75	0.93	91.78	71.76	61.5	61.51	-32.98%	-32.99%
1.00	1.00	1.00	0.45	1.49	78.34	19.97	29.14	27.02	-65.51%	-62.80%
1.00	1.00	0.50	1.00	1.31	116.53	28.43	28.18	27.2	-76.66%	-75.82%
1.50	1.50	0.50	1.50	1.95	29.87	8.54	8.65	6.57	-78.00%	-71.04%
2.00	1.20	1.50	0.50	2.47	11.66	5.33	7.15	5.43	-53.43%	-38.68%
2.00	1.50	1.80	2.00	2.67	8.45	3.34	2.83	3.58	-57.63%	-66.51%

Comparación de rendimientos utilizando cuatro variables. Escenario (i), los coeficientes de las componentes principales son  $\alpha_{1a} = -0.52$ ,  $\alpha_{2a} = -0.50$ ,  $\alpha_{3a} = -0.37$ ,  $\alpha_{4a} = -0.58$ ;  $\alpha_{1b} = -0.26$ ,  $\alpha_{2b} = 0.49$ ,  $\alpha_{3b} = -0.77$ ,  $\alpha_{4b} = 0.29$ ;  $\alpha_{1c} = 0.70$ ,  $\alpha_{2c} = -0.49$ ,  $\alpha_{3c} = -0.50$ ,  $\alpha_{4c} = 0.11$ ;  $\alpha_{1d} = 0.41$ ,  $\alpha_{2d} = 0.51$ ,  $\alpha_{3d} = -0.09$ ,  $\alpha_{4d} = -0.75$ , y sus varianzas son  $\lambda_a = 1.95$ ,  $\lambda_b = 0.93$ ,  $\lambda_c = 0.68$ ,  $\lambda_d = 0.44$ .