

**ESCUELA POLITÉCNICA NACIONAL**

FACULTAD DE CIENCIAS

DISEÑO DEL GRÁFICO DE CONTROL NO PARAMÉTRICO BASADO  
EN LA DISTANCIA DE MAHALANOBIS GENERALIZADA PARA  
DATOS FUNCIONALES

TRABAJO DE TITULACIÓN PREVIO A LA OBTENCIÓN DEL TÍTULO  
DE INGENIERA MATEMÁTICA

PROYECTO DE INVESTIGACIÓN

Priscila Salomé Guayasamín Bahamonde  
priscillaguayasamin@gmail.com

Director: Ph.D. Miguel Alfonso Flores Sánchez  
miguel.flores@epn.edu.ec

QUITO, MARZO 2022

# DECLARACIÓN

Yo PRISCILA SALOMÉ GUAYASAMÍN BAHAMONDE, declaro bajo juramento que el trabajo aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su Reglamento y por la normatividad institucional vigente.



Firmado electrónicamente por:

**PRISCILA SALOME  
GUAYASAMIN  
BAHAMONDE**

---

Priscila Salomé Guayasamín Bahamonde

# CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por Priscila Salomé Guayasamín Bahamonde, bajo mi supervisión.

---

Ph.D. Miguel Alfonso Flores Sánchez

# AGRADECIMIENTOS

A Dios por todas sus bendiciones.

A Miguel, mi tutor, por su tiempo y paciencia al dirigir este trabajo.

A José y Teresa, mis padres, por su apoyo incondicional.

A Pamela, Alexis y Onella, mis hermanos, por contagiarme su alegría.

A Sebastián, mi cómplice, por caminar junto a mí en todo momento.

A Angelita, mi familia, por su corazón generoso.

A Paola, Cristian, Gissela, Maru, mis amigos, por su lealtad.

# DEDICATORIA

*A Manuel y Ángela*

# Contenidos

<b>Contenidos</b>	<b>2</b>
<b>Tablas</b>	<b>5</b>
<b>1. Introducción</b>	<b>10</b>
1.1. Planteamiento del problema . . . . .	10
1.2. Objetivos . . . . .	12
1.2.1. Objetivo General . . . . .	12
1.2.2. Objetivos Específicos . . . . .	12
<b>2. Control estadístico de procesos</b>	<b>13</b>
2.1. Control, proceso y calidad . . . . .	14
2.2. Metodología Seis Sigma . . . . .	15
2.3. Gráficos de Control . . . . .	17
2.3.1. Elección de los límites control . . . . .	21
2.4. Relación entre las cartas de control y las pruebas de hipótesis .	22
2.5. Rendimiento de Gráficos de Control . . . . .	23
2.6. Tipos de gráficos de control . . . . .	24
2.7. Monitorizado de bases de datos de alta dimensión . . . . .	25
<b>3. Análisis de datos Funcionales</b>	<b>27</b>

---

3.1.	Espacios de Hilbert separables . . . . .	28
3.2.	Integral vectorial . . . . .	29
3.2.1.	Integral de Bochner . . . . .	29
3.3.	Elementos aleatorios en espacios de Hilbert . . . . .	31
3.3.1.	Medidas de probabilidad en un espacio de Hilbert . . . . .	31
3.3.2.	Media y covarianza de un elemento aleatorio en un espacio de Hilbert . . . . .	32
3.4.	Estadístico basado en la distancia de Mahalanobis: $T^2$ de Hotelling	34
3.4.1.	Estadístico $T^2$ en el espacio de Hilbert separable . . . . .	35
3.4.2.	Cálculo numérico del estadístico $T^2$ de Hotelling . . . . .	36
3.4.3.	Estadístico basado en la distancia de Mahalanobis a dos muestras . . . . .	38
3.4.4.	Distancias para datos funcionales . . . . .	38
3.5.	Técnicas de remuestreo bootstrap . . . . .	39
3.5.1.	Bootstrap suavizado . . . . .	41
3.5.2.	Permutaciones . . . . .	43
3.6.	Medidas de Profundidad . . . . .	44
<b>4.</b>	<b>Propuesta metodológica para la fase de monitorización</b>	<b>48</b>
4.1.	Metodología . . . . .	49
4.2.	Procedimiento de construcción de un gráfico de control para la fase de monitorización . . . . .	50
4.3.	Estudio de Simulación . . . . .	52
4.3.1.	Medición del rendimiento del gráfico de monitorización . . . . .	55
4.4.	Caso de Estudio: Aneurismas de pacientes de alto y bajo riesgo	78

<b>5. Conclusiones y líneas futuras</b>	<b>84</b>
5.1. Conclusiones y líneas futuras . . . . .	84
5.1.1. Conclusiones . . . . .	84
5.1.2. Líneas Futuras . . . . .	85
<b>A. Anexos: Código</b>	<b>87</b>
<b>B. Anexos: Espacios de vectores y funciones</b>	<b>103</b>
B.0.1. Espacios vectoriales y normados . . . . .	105
B.0.2. Espacios de Hilbert y dotados de un producto inercial . .	106
B.0.3. Operadores . . . . .	108
B.0.4. Operadores Adjuntos . . . . .	110
B.0.5. Operadores Compactos . . . . .	110
B.0.6. Operadores de Hilbert Schmidt . . . . .	111
<b>Bibliografía</b>	<b>117</b>



# Tablas

2.1. Tipos de cartas por variables . . . . .	24
2.2. Tipos de cartas por atributo . . . . .	24
4.1. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 1A para $n=25$ y $m=25$ . . . . .	61
4.2. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 1A para $n = 25$ y $m = 50$ . . . . .	62
4.3. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 1A para $n = 25$ y $m = 100$ . . . . .	63
4.4. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 1A para $n = 50$ y $m = 50$ . . . . .	64
4.5. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 1A para $n = 50$ y $m = 100$ . . . . .	65
4.6. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1-\beta$ en el escenario 1A para $n = 100$ y $m = 100$ . . . . .	66

---

4.7. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 1B para $n = 25$ y $m = 25$ . . . . .	67
4.8. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 1B para $n = 25$ y $m = 50$ . . . . .	68
4.9. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 1B para $n = 25$ y $m = 100$ . . . . .	69
4.10. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 1B para $n = 50$ y $m = 50$ . . . . .	70
4.11. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 1B para $n = 50$ y $m = 100$ . . . . .	71
4.12. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 1B para $n = 100$ y $m = 100$ . . . . .	72
4.13. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 2A. Gráfico basado en la distancia de Mahalanobis . . . . .	73
4.14. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 2A. Gráfico basado en el estadístico $L^1$ . . . . .	74
4.15. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 2A. Gráfico basado en el estadístico $L^2$ . . . . .	75

---

4.16. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 2B. Gráfico basado en la distancia de Mahalanobis . . . . .	76
4.17. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 2B. Gráfico basado en el estadístico $L^1$ estandarizado . . . . .	77
4.18. Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control $1 - \beta$ en el escenario 2B. Gráfico basado en el estadístico $L^2$ estandarizado . . . . .	78

# Resumen

El gráfico de control es una de las herramientas principales en el control estadístico de procesos que permiten recopilar y analizar datos con la finalidad de evaluar, monitorear y controlar el rendimiento de procesos en general. A través del uso de los gráficos de control se puede identificar variaciones en el proceso debido a causas inherentes y causas especiales.

En el presente trabajo se propone el gráfico de control basado en la distancia de Mahalanobis extendido para datos funcionales para el monitoreo de procesos, donde las características a ser monitoreadas pueden ser vistas como funciones suaves. En la literatura actual se han propuesto gráficos multivariantes paramétricos; sin embargo, en este trabajo se diseñan gráficos de control no paramétricos  $L^1$  estandarizado,  $L^2$  estandarizado, y la carta de control basada en la distancia de Mahalanobis, considerando los procedimientos de bootstrap, permutaciones y medidas de profundidad. Se plantea un estudio de simulación donde se compara el rendimiento de las cartas de control considerando el tamaño de las muestras y el método de determinación del umbral que discrimina las curvas atípicas.

Finalmente, se aplica la metodología desarrollada para el caso de estudio de aneurismas cerebrales de pacientes de alto y bajo riesgo, donde se monitorizará la derivada del radio de los aneurismas a lo largo de la arteria carótida interna para determinar el efecto de un procedimiento médico en los pacientes en el tiempo.

**Palabras clave:** control estadístico de procesos, gráficos de control no paramétricos, bootstrap, permutaciones, Mahalanobis, medidas de profundidad.

# Abstract

Control charts are a fundamental tool in statistical process control, since they allow the gathering and analysis of data in order to evaluate, monitor and control the performance of processes. Furthermore, control charts are helpful in identifying variations in processes due to common and special causes. In this paper, we propose a Mahalanobis distance-based control chart with functional data for monitoring processes presenting features that can be viewed as smooth functions. In current literature, multivariate parametric graphs have been proposed in order to monitor such processes. However, in the present work we design a nonparametric control chart which is based on the Mahalanobis distance besides nonparametric control charts based on standardized  $L^1$  and standardized  $L^2$  statistics. All control charts were designed using bootstrap methods, permutations and depth measurements. Additionally, we present a simulation study where control chart performance comparisons are made using different sample sizes and threshold determining methods which detect atypical curves. Finally, the methodology developed in this work is used to analyze the effect of a medical procedure on patients with high and low risk of brain aneurysm. In particular, our method is used to monitor the time derivative of the radius of aneurysms along the internal carotid artery.

**Key words:** statistical process control, nonparametric control charts, bootstrap, permutations, Mahalanobis, depth measurements

## 1.1 Planteamiento del problema

Los datos funcionales surgen cuando las variables de interés se pueden ver de forma natural como una curva o una función suave. Por ende, el Análisis de Datos Funcionales (FDA) se puede considerar como un análisis estadístico de muestras de curvas. Varias técnicas estadísticas clásicas han sido extendidas al análisis funcional. No obstante, en el área del control de la calidad, la monitorización de variables de carácter funcional es un problema en investigación [1].

Dentro del control estadístico de la calidad se presentan varias herramientas y técnicas, siendo una de las más usuales la carta o gráfico de control. En 1920, el Dr. Walter Shewart [2] introdujo por primera vez el gráfico de control como una herramienta útil para detectar anomalías y estabilizar un proceso. Se dice que un proceso es estable o está bajo control cuando la única variabilidad existente se debe a causas inherentes del mismo. Si se asume que el proceso es estable la construcción del gráfico de control se centra en el diseño de la fase de monitorización la cual se encarga de controlar nuevas observaciones.

Para el caso univariante, las cartas de control monitorizan procesos de una sola variable. Sin embargo, en la práctica la mayoría de los escenarios de monitoreo incluyen algunas variables relacionadas, con lo cual surgen los gráficos de control multivariante. En 1947, Harold Hotelling [3] propuso un gráfico de control multivariante basado en la distancia de Mahalanobis; la cual mide la distancia entre la media de un proceso bajo control y una observación; o, la media del proceso bajo control y la media de un conjunto de nuevas observaciones. Así también, las distancias

estandarizadas  $L^1$  y  $L^2$  de acuerdo a Hall and Tajvidi [4] han sido ampliamente usadas.

En la fase de monitorización, a partir de la distribución del estadístico asociado al gráfico de control, se requiere determinar el cuantil a partir del cual una o más observaciones se consideran atípicas. En el campo paramétrico, este problema está solventado, pues el estadístico basado en la distancia de Mahalanobis tiene una distribución conocida, bajo el supuesto (fuerte) de normalidad. En cambio, en el campo no paramétrico en el contexto multivariante, por lo general, para determinar el límite de control superior se aplica la técnica de remuestreo bootstrap clásico a las observaciones [5].

En el contexto funcional, en el área del control estadístico de la calidad, uno de los problemas que se enfrenta es el diseño de gráficos de control para datos funcionales. Los nuevos datos que caracterizan los procesos industriales actuales requieren el desarrollo de soluciones innovadoras a partir de la aplicación de técnicas de análisis de datos funcionales (FDA). Entre las investigaciones más destacadas en este contexto, Colosimo y Pacella [6] realizan un estudio de comparación del desempeño de diferentes gráficos de control para monitorizar datos funcionales. Por otro lado, Sheu et al.[7] propone un gráfico de control para la monitorización de datos funcionales basado en remuestreo bootstrap, mientras que García [8] y Rodrigo [9] se adaptan los conceptos de gráficos de control al contexto de los datos funcionales.

Por lo mencionado anteriormente y, debido a que en el campo multivariante la carta de control basada en distancia de Mahalanobis ha sido ampliamente usada, se busca entonces, diseñar la fase de monitorización del gráfico de control no paramétrico basado en la distancia de Mahalanobis para contrastarlo con las cartas de control  $L^1$  y  $L^2$  estandarizadas, aplicando procedimientos alternativos como bootstrap y permutaciones, con la finalidad de determinar la distribución del estadístico y con ello el umbral que detecte las anomalías de variables funcionales.

## 1.2 Objetivos

### 1.2.1 Objetivo General

Desarrollar la metodología para el diseño de la fase de monitorización del gráfico de control no paramétrico, basado en la distancia de Mahalanobis,  $L^1$  y  $L^2$  estandarizadas para datos funcionales.

### 1.2.2 Objetivos Específicos

- Extender conceptos y técnicas estadísticas multivariantes al caso funcional necesarias para el diseño de fase de monitorización del gráfico de control no paramétrico basado en la distancia de Mahalanobis,  $L^1$  y  $L^2$  estandarizadas.
- Definir escenarios de simulación para comparar el desempeño del gráfico de control basado en la distancia de Mahalanobis con las cartas de control  $L^1$  y  $L^2$  estandarizadas, considerando diferentes tamaños de muestra estable, de monitoreo, procedimiento bootstrap y permutaciones.
- Programar en el software estadístico R el algoritmo de la fase de monitoreo de los gráficos de control.
- Aplicar la metodología establecida para la monitorización de la derivada del radio del aneurisma en función la longitud de curva de la ACI, comparando la media de los pacientes de alto riesgo con los de bajo riesgo.



## CAPÍTULO 2

El Control de la Calidad es un concepto originalmente desarrollado en el ámbito de la producción, y en general, a través de los años se ha consolidado como un conjunto de técnicas estadísticas para procesos, con el objeto de comprobar si todas o cada parte de un producto o servicio cumplen ciertas exigencias de calidad [10].

Considerando la calidad de un producto, Seis Sigma es una metodología que desarrolla la mejora continua de procesos, pues se centra en reducir y eliminar los defectos o fallos. Dentro de este modelo de gestión de calidad existe una etapa importante denominada control, donde se aplican las medidas necesarias que garanticen la eficacia y continuidad del proceso. Una de las herramientas potentes utilizadas en dicha etapa es el gráfico de control. En este sentido, la carta de control es una de las técnica más usadas para mejorar la calidad de los productos. En este trabajo se abordan los gráficos de control no solo como una herramienta enfocada a la producción sino, en general como una técnica para la detección de anomalías[11].

Hace aproximadamente un siglo se inició un nuevo campo de aplicación de la estadística impulsado por el Dr. W. Shewhart. Este físico, ingeniero y estadístico norteamericano es considerado el padre del control estadístico de la calidad. Sus teorías sobre el control de procesos y en especial los gráficos de control propuestos por él, son la base que ahora se conoce como Control Estadístico de Procesos. En 1974, el Dr. Kaoru Ishikawa incluye al gráfico de control como una de las siete principales herramientas del control estadístico de la calidad. A lo largo del tiempo se han ido desarrollando cartas univariadas para variables continuas o discretas, gráficos de control que guardan información de eventos pasados, gráficos de control multivariantes y sus variantes más importantes. En esta tesis nos enfocaremos en la fase de monitorización de una carta de control que es importante porque permite

detectar posibles anomalías a lo largo del tiempo, esta fase hace posible identificar situaciones fuera de control para corregirlas inmediatamente e implementar mejoras.

En 1947 el profesor H. Hotelling, propuso un gráfico de control multivariado basado en la distancia de Mahalanobis, hoy conocido como gráfico  $T^2$  de Hotelling y que es el equivalente al gráfico de Shewhart en el caso univariado, esta carta de control combina información de la media y dispersión de más de una característica para reducir la variabilidad [12]. En el trabajo de Mason y Young [13] se presenta un amplio desarrollo relacionado con la metodología para el diseño de dicho gráfico de control y sus variantes de acuerdo al número de observaciones. Además, podemos encontrar amplias utilidades en el campo de la industria [13]. En este capítulo se presenta la carta de control basada en la distancia de Mahalanobis desde la perspectiva multivariante, que más adelante se extenderá al contexto de espacios de dimensión infinita.

El análisis estadístico de datos en espacios infinitos dimensionales es un área activa de investigación. En el caso particular del control estadístico de la calidad surge la necesidad de desarrollar técnicas y metodologías que extiendan los conceptos multivariantes al contexto funcional. Es decir, que los datos puedan ser representados como funciones continuas, lo que amplía la teoría de control de calidad del campo multivariante al funcional. Así, lo mencionado en el párrafo anterior y en este, constituyen una motivación para el desarrollo del presente trabajo.

## 2.1 Control, proceso y calidad

El *control estadístico de procesos* (CEP ó SCP por sus siglas en inglés) se define como un conjunto de técnicas estadísticas para controlar un proceso o método de producción. En general, las herramientas y procedimientos de CEP ayudan a monitorear el comportamiento del proceso para desarrollar un mejoramiento continuo de la calidad [10]. En el libro Control Estadístico de la Calidad de Montgomery [11] se puede encontrar una revisión más detallada sobre el desarrollo histórico de la metodología del mejoramiento de calidad y una visión general de las herramientas

estadísticas esenciales para el control estadístico.

La idea de control se relaciona con la gestión de un *proceso* o la gestión de la calidad. De acuerdo a [10] un proceso es un conjunto de actividades entrelazadas o interrelacionadas que reciben determinados insumos (entradas) que son transformados en un resultado (salidas) o en un producto. Las variables de salida son características de calidad o variables de respuesta, en las que se reflejan los resultados obtenidos en el proceso.

En lo referente a la *calidad*, existen muchas maneras de definirla, donde el concepto más extendido se relaciona con las características deseables que debería tener un producto o servicio. Por ejemplo, se puede definir la calidad como la adecuación para el uso, la cual es inversamente proporcional a la variabilidad, y por ende, el objetivo principal es intentar reducir la variabilidad en los procesos [14].

En cualquier proceso, independientemente de lo adecuado que sea su diseño, siempre existirá una cierta cantidad de variabilidad. En este sentido, la metodología Seis Sigma desarrolla la mejora continua de productos o servicios reduciendo y eliminando defectos. En la siguiente sección se describirá brevemente dicha metodología.

## 2.2 Metodología Seis Sigma

Seis Sigma es una metodología rigurosa que incorpora un enfoque sistemático de la reducción de variabilidad en un proceso. A partir de herramientas estadísticas busca determinar y analizar los posibles problemas que afecte un proceso.

La metodología Seis Sigma utiliza un sistema de indicadores para diagnosticar el estado de un proceso así como para medir los resultados de un proyecto de mejora. Los índices más frecuentes son el índice de defectos por unidad y el índice de defectos por oportunidad. Estos indicadores permiten establecer comparaciones entre procesos.

$$DPU = \frac{Defectos}{Unidades}$$

$$DPO = \frac{Defectos}{Unidades \times Oportunidad}$$

El *DPU* es el número de defectos en una muestra dividido entre el número de unidades de la muestra. El *DPO* es el número de defectos en una muestra dividido entre el número total de oportunidades de defectos. También se suele hablar del índice de defectos por millón de oportunidades (DPMO) que no es más que el DPO multiplicado por un millón.

Los indicadores anteriores dependen de la variabilidad del proceso que se analiza. Mientras mayor dispersión tienen las variables que se estudian, mayor es la proporción de unidades fuera de las especificaciones.

Si la variable de calidad de interés en un proceso de manufactura o de servicio tiene distribución normal, el intervalo comprendido entre la media  $+/-$  tres veces la desviación estándar contiene aproximadamente el 99,73 % de las unidades, si dichos límites coinciden con las especificaciones, el proceso tendrá 0,27 % de unidades defectuosas o 2700 unidades por millón. Un proceso con estas características es aceptable para mejorar la calidad, sin embargo, en la metodología Seis Sigma el objetivo es obtener tasas de defectos más pequeñas. Se busca reducir la variabilidad del proceso de modo que la media llegue a desplazarse hasta 1,5 veces la desviación estándar (en más o en menos), así quedarán fuera 3,4 unidades por millón. A esta clase de procesos se los llama Seis Sigma.

Puede notarse que aceptar un desplazamiento del valor medio en 1,5 veces la desviación estándar, equivale a obtener una proporción de unidades fuera de especificación igual al área encerrada por la curva normal a la derecha del valor medio más 4,5 veces la desviación estándar (o a la izquierda del valor medio menos

4,5 veces la desviación estándar ). Esta área corresponde a 0,00034 % o su equivalente 3,4 unidades por millón.

Cualquier proceso dentro de una organización puede entenderse como una función que aplicada sobre ciertas variables de entrada, proporciona un conjunto de variables resultado. La metodología Seis Sigma mejora los procesos sobre las variables de entrada que influyen significativamente sobre las variables de resultados.

La metodología Seis Sigma se caracteriza por un proceso estructurado de cinco etapas concretas:

- La fase Definir consiste en realizar un diagnóstico para identificar cuáles son los elementos que participan en el proceso.
- En la etapa Medir se deben determinar cuáles las características críticas que influyen sobre las variables resultantes del proceso y medirlas.
- En la fase Analizar se estudia exhaustivamente toda la información recolectada en la etapa anterior para identificar las causas de la variación del proceso.
- La fase Mejorar comienza una vez que se han identificado las causas. En esta etapa se desarrolla un plan de mejoras para eliminar los defectos en los que incurre el proceso.
- El objetivo de la fase Controlar es el completar y monitorear los resultados obtenidos en las etapas anteriores, así como documentar nuevas especificaciones del proceso de estudio. La técnica más representativa en la etapa de controlar es el gráfico de control que se presenta en la siguiente sección.

## 2.3 Gráficos de Control

Un gráfico de control es una técnica de monitoreo cuyo objetivo es observar y analizar el comportamiento de un proceso a través del tiempo. La carta de control de Shewhart

fue desarrollada en los años 1920 por el doctor Walter A. Shewhar de Bell Telephone Laboratories, así, una de las primeras definiciones a entender es la variabilidad de Shewhart [10].

En cualquier proceso, independientemente de lo adecuado que sea su diseño o de la atención que se preste a su mantenimiento, siempre existirá cierta cantidad de variabilidad inherente o natural. Esta variabilidad natural o “ruido de fondo”, es el efecto acumulado de muchas causas pequeñas y en esencia inevitables. En el contexto del control estadístico de calidad, a esta variabilidad natural se le denomina un “sistema estable de causas fortuitas”. Se dice que un proceso que opera únicamente con causas fortuitas de variación está bajo control estadístico. En otras palabras, las causas fortuitas son parte inherente del proceso.

En ocasiones puede estar presente otra clase de variabilidad en la salida de un proceso. Esta variabilidad en las características de la calidad se originan de tres fuentes: máquinas ajustadas o controladas incorrectamente, errores del operador o materia prima defectuosa. En general, esta variabilidad es grande cuando se la compara con el ruido de fondo, y suele representar un nivel inaceptable del desempeño del proceso. A estas fuentes de variabilidad que no son parte del patrón de las causas fortuitas se las llama “causas asignables”. Se dice que un proceso que opera en presencia de causas asignables está fuera de control.

Con frecuencia, los procesos de producción operarán en el estado bajo control, produciendo productos aceptables durante períodos relativamente prolongados. Con el tiempo, sin embargo, ocurrirán causas asignables, aparentemente al azar, que ocasionarán un “corrimiento” a un estado fuera de control en el que una proporción mayor de la salida del proceso no cumplirá con los requerimientos. Uno de los objetivos principales del control estadístico de procesos es detectar con rapidez la ocurrencia de causas asignables en el corrimiento del proceso a fin de hacer la investigación pertinente y emprender las acciones correctivas antes de que se fabriquen muchas unidades defectuosas. La carta de control es una técnica de monitoreo de procesos en línea que se usa ampliamente para este fin, que puede ofrecer información útil para

mejorar el proceso.

La idea principal de un gráfico de control es que, a través del cálculo de límites de control, podamos observar dónde varía el proceso a través del tiempo, graficando un estadístico denotado por  $W$  el cual mide la característica de interés en el proceso. Los elementos para construir una carta de control son:

- Línea central (LC): Esta línea representa el promedio de los valores de  $W$ . Es la media teórica de las observaciones de la muestra.
- Límites de control inferior (LCI) y superior (LCS): Estos límites definen el rango de variación del proceso, de forma que al estar el proceso bajo control, hay una alta probabilidad de que los valores de  $W$  se encuentren dentro de los límites de control. Estos límites son simétricos respecto la línea central si la distribución de probabilidad de la variable es simétrica.

Los límites de control inferior y superior son calculados considerando la variación de los datos que se presentan en la carta. Se debe tener mucho cuidado en la elección de los límites porque si se desea cubrir un alto porcentaje de variabilidad los límites serán muy amplios, lo que dificultaría la detección de atípicos, por otra parte, si el porcentaje es pequeño los límites serán estrechos, lo que puede causar muchas señales en falso.

De acuerdo a Montgomery, en [10], se presentan las siguientes utilidades de los gráficos de control.

- Las cartas de control son una técnica probada para mejorar la productividad. Un programa exitoso de cartas de control reducirá los desechos y el reprocesamiento, que son los principales depredadores de la productividad en cualquier operación. Si se reducen los desechos y el reprocesamiento se incrementará la productividad, se reducirán los costos y se aumentará la capacidad de producción.

- Las cartas de control son efectivas para prevenir defectos. La carta de control ayuda a mantener el proceso bajo control, lo cual es consistente con la filosofía del “hacerlo bien a la primera”.
- Las cartas de control previenen el ajuste innecesario del proceso. Una carta de control puede distinguir el ruido de fondo de la variación anormal. Si los operadores del proceso ajustan el proceso con base en pruebas periódicas que no guarden relación con un programa de cartas de control, con frecuencia tendrán una reacción exagerada al ruido de fondo y realizarán ajustes innecesarios.
- Las cartas de control proporcionan información sobre la capacidad del proceso. La carta de control proporciona información del valor de parámetros importantes del proceso y de su estabilidad con el tiempo. Dicha información es de gran utilidad para los diseñadores del proceso.

En la siguiente figura se ilustra un gráfico de control con los límites descritos anteriormente, donde se puede observar como varía determinado proceso, es decir, observaciones bajo y fuera de control.

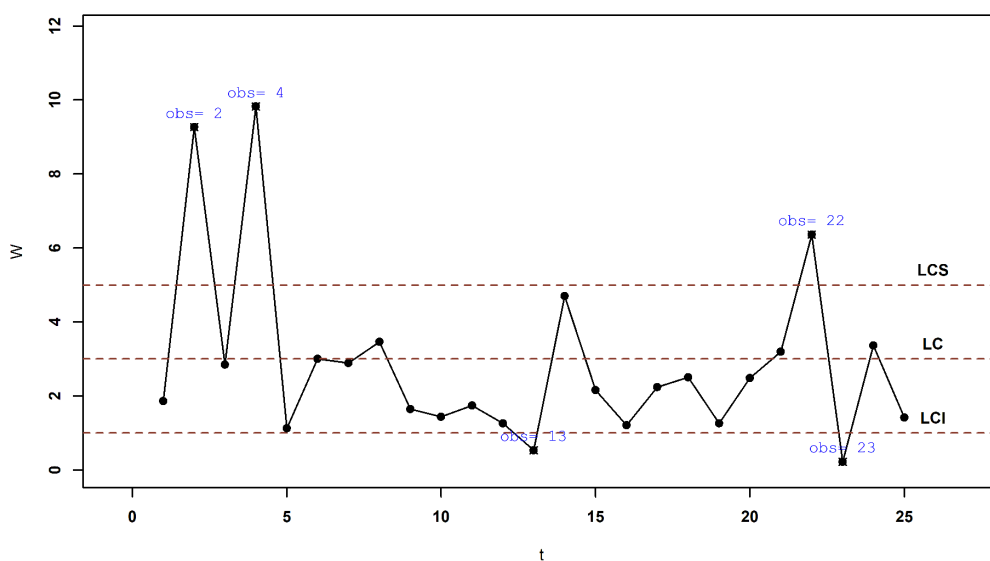


Figura 2.1: Ejemplo de un gráfico de control



### 2.3.1 Elección de los límites control

La especificación de los límites de control es una de las decisiones críticas que se deben tomar al diseñar una carta de control. Estos límites de control se eligen de tal modo que si el proceso está bajo control, casi todos los puntos muestrales estarán entre ellos. Sin embargo, un punto que se localiza fuera de los límites de control se interpreta como evidencia de que el proceso está fuera de control, lo que requiere investigación y acción correctiva para encontrar y eliminar la causa o causas asignables responsables de este comportamiento.

La manera más frecuente de elegir estos límites es a partir de la relación entre la media y las desviación estándar del estadístico  $W$  que se monitorea. En el caso de que  $W$  siga una distribución normal con media  $\mu_W$  y desviación estándar  $\sigma_W$ , se tiene que los límites están dados por  $\mu_W - 3\sigma_W$  y  $\mu_W + 3\sigma_W$ , donde bajo control estadístico se ubica el 99.73 % de los posibles valores de  $W$ .

En el caso de tener una distribución diferente a la normal, se aplica extensión del teorema de Chebyshev que se enuncia a continuación.

**Teorema 2.3.1.** (*Desigualdad de Chebyshev*) Sea  $\mu = E(X)$  y  $\sigma^2 = Var(X)$ , entonces  $P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$

A partir de ese resultado se obtiene una relación entre  $\bar{X}$  y  $S$  que establece el porcentaje mínimo de datos que caen entre los límites  $\bar{X} - kS$  y  $\bar{X} + kS$ . Las cartas de control que siguen esta metodología fueron propuestas por Walter A. Shewhar, y sus límites de control están dados por

$$\begin{aligned} LCI &= \mu_W - 3\sigma_W \\ LC &= \mu_W \\ LCS &= \mu_W + 3\sigma_W \end{aligned} \tag{2.1}$$

Este tipo de cartas son conocidas como cartas tipo Shewhart. Con ellas y bajo condiciones del control estadístico se tendrá alta probabilidad de que los valores de  $W$  caigan dentro de los límites definidos en 2.1.

## 2.4 Relación entre las cartas de control y las pruebas de hipótesis

Existe una estrecha relación entre las cartas de control y la prueba de hipótesis. En cierto sentido, la carta de control es una prueba de hipótesis de que el proceso está en un estado de control estadístico. Un punto que se localiza dentro de los límites de control es equivalente a no poder rechazar la hipótesis de control estadístico, y un punto que se localiza fuera de los límites de control es equivalente a rechazar la hipótesis de control estadístico.

Una situación que se presta para el marco de la prueba de hipótesis es el análisis del desempeño de un gráfico de control. Por ejemplo, podría considerarse la probabilidad del error tipo I de la carta de control (concluir que el proceso está fuera de control cuando en realidad está bajo control) y la probabilidad del error tipo II de la carta de control (concluir que el proceso está bajo control cuando en realidad está fuera de control).

Cuando los límites de control se alejan más de la línea central, se reduce el riesgo de un error tipo II, es decir, el riesgo de que un punto se localice fuera de los límites de control, indicando una condición fuera de control cuando no está presente ninguna causa asignable.

Sin embargo, al ensanchar los límites de control se incrementará también el riesgo del tipo II, es decir, al riesgo de un punto se localice dentro de los límites de control cuando el procesos en realidad está fuera de control.

	$H_0$ es verdad	$H_1$ es verdad
No se rechaza $H_0$	Decisión Correcta $1 - \alpha$	Error tipo II $\beta$
Se rechaza $H_0$	Error Tipo I $\alpha$	Decisión Correcta $1 - \beta$

Tipos de error

## 2.5 Rendimiento de Gráficos de Control

Para medir el rendimiento del gráfico de control usamos el promedio de longitud de rachas (ARL). Es decir, el número de muestras sucedidas hasta obtener una señal de fuera de control.

En vista de que  $H_0$  es un estado bajo control del proceso y que la probabilidad de rechazar que el proceso está bajo control, es  $\alpha$ , entonces la medida utilizada de falsas alarmas es  $ARL_0 = \frac{1}{\alpha}$ . El  $ARL_0$ , depende solo de  $\alpha$ , cuyo valor lo determina el investigador.

De la misma forma,  $H_1$  es la hipótesis que el proceso no está bajo control por lo que, la probabilidad de no rechazar que el proceso está bajo control cuando no lo está, es  $\beta$ , entonces una medida utilizada de un estado fuera de control es  $ARL_1 = \frac{1}{1-\beta}$ . En este caso lo que se obtendría es el inverso de la potencia de la prueba.

Está claro que se desea que las falsas alarmas de que el proceso está fuera de control sean menos frecuentes. por ejemplo a menor  $\alpha$  mayor  $ARL_0$ . Así mismo, necesitamos detectar pronto cuando el proceso está fuera de control, por eso necesitamos un  $\beta$  pequeño, lo que da como resultado un  $ARL_1$  pequeño.

## 2.6 Tipos de gráficos de control

El control de la calidad mediante el término variable designa a cualquier característica de calidad como medible tal como el tiempo, el peso, la temperatura, entre otros. Mientras que se denominan atributo a las características de calidad que no se miden y pueden ser conforme y disconforme o defectuoso y no defectuoso.

Según sea el tipo de característica de calidad (variable o atributo) a controlar será la carta de control correspondiente. Por lo tanto, las cartas de control se clasifican en cartas de control por variables y cartas de control por atributos.

En las siguientes tablas se presentan los principales gráficos de control y sus principales campos de aplicación. Cabe recalcar que la aplicación de estos gráficos de control se basa en independencia y en una distribución paramétrica de los datos.

Carta	Descripción	Campo de Aplicación
$\bar{X}$ -R	Medias y Rangos	Control de características individuales
$\bar{X}$ -S	Medias y desviación estándar	Control de características individuales
I-MR	Individuales	Control de un proceso con datos que no pueden ser muestreados en lotes o grupos

Tabla 2.1: Tipos de cartas por variables

Carta	Descripción	Campo de Aplicación
P	Proporciones o fracción defectiva	Control de la fracción global de defectuosos de un proceso
NP	Número de defectuosos	Control del número de piezas defectuosas
C	Defectos por unidad	Control de número global de defectos por unidad
U	Promedio de defectos por unidad	Control del promedio de defectos por unidad

Tabla 2.2: Tipos de cartas por atributo

En los procesos de producción actuales, para controlar la calidad de los productos es necesario vigilar más de una característica simultáneamente. Una primera posibilidad sería desarrollar métodos estadísticos de control para cada una de las características, es decir, aplicar técnicas de control de calidad univariantes de forma independiente. Sin embargo, sería imposible controlar el efecto de las interacciones entre las diversas variables así como su variación a lo largo del tiempo: en términos estadísticos, se estaría obviando la información contenida en las covarianzas. Con el fin de controlar varias variables de calidad con su estructura de correlación, se desarrollan 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. Otros estadísticos mutivariantes que consideran la correlación de un conjunto de variables son los propuestos por Crossier en [15] y Lowry et al. en [16].

En la literatura se han propuesto una gran cantidad de gráficos de control multivariantes, sin embargo, la mayoría de ellos se construyen con el mismo procedimiento del gráfico basado en la distancia de Mahalanobis o sus variaciones. A continuación se expone en forma general los gráficos de control que monitorizan la media desde en el campo univariante y multivariante.

## **2.7 Monitorizado de bases de datos de alta dimensión**

En los últimos años la calidad de variables monitorizadas y su complejidad han crecido exponencialmente [17]. Últimamente se está monitoreando dichas bases de datos teniendo en cuenta la autocorrelación existente y efectos estacionales. En la actualidad se sugiere la aplicación del gráfico de control basado en la distancia de Mahalanobis, técnicas de clasificación supervisada, regresión multivariante, series de tiempo multivariadas, combinadas con técnicas de reducción de dimensiones como análisis de componentes principales, análisis factorial y mínimos cuadrados parciales [19], [20], [21]. Por lo mencionado anteriormente, el monitorizado de bases de datos de alta dimensión requiere de una labor investigativa que permita en general la detección de anomalías en tiempo real.

Dentro del análisis multivariante, se encuentra en investigación el control espacio-temporal y el control de perfiles. En lo correspondiente al primero, dentro del control estadístico de procesos se añade las variaciones espaciales y se encuentran interesantes contribuciones de Well et al. [22], donde se propone la monitorización a partir de imágenes.

De acuerdo a Megaged y Jones Farmer [23], el control de perfiles puede afrontarse desde la perspectiva de datos funcionales. La monitorización de perfiles es un conjunto de técnicas estadísticas aplicadas cuando la calidad de un proceso se mide por la relación funcional entre dos variables. Se suelen involucrar el ajuste de modelos de regresión y la monitorización mediante el gráfico de control basado en la distancia de Mahalanobis y otras alternativas multivariantes.

## CAPÍTULO 3

El análisis de datos en espacios de dimensión infinita, datos funcionales, es una área activa en la estadística. En los trabajos de Ramsay & Silverman [24] y de Ferraty & Vieu [1], se muestran muchas ventajas del análisis de datos funcionales, las cuales son cada vez más populares en estas últimas décadas. El hecho de evaluar los datos en tiempo continuo, es decir, considerarlos funciones aleatorias de un proceso estocástico, ha hecho que surjan problemas teóricos y prácticos, por lo cual se ve la necesidad de extender conceptos del campo multivariante al funcional, para proponer nuevas técnicas y metodologías.

En este capítulo se presentarán los insumos teóricos para la construcción de la fase de monitorización del gráfico de control basado en la distancia de Mahalanobis,  $L^1$  y  $L^2$  estandarizadas. Para esto, en la literatura se presentan dos macro metodologías para definir dicho estadístico. El primer enfoque se refiere a proyectar las curvas funcionales en un subespacio de menor dimensión usando una expansión de base truncada para evaluar el estadístico en dicho subespacio. Este desarrollo se presenta en los trabajos de Spitzner et al. [25], Cuesta-Albertos y FebreroBande [26]. El segundo enfoque consiste en analizar los datos en su espacio original. En este trabajo, se presentará la re definición del estadístico basado en la distancia de Mahalanobis ( $T^2$  de Hotelling) en el espacio separable de Hilbert propuesta en [27]. Es decir, se seguirá el segundo enfoque para evitar obviar la información que se pierde al proyectar los datos en subespacios de menor dimensión. Adicionalmente, se presentarán las definiciones de las distancias estandarizadas  $L^1$  y  $L^2$  en el contexto funcional, que de acuerdo a Hall and Tajvidi [4] han sido ampliamente usadas con la finalidad de comparar el desempeño de la fase de monitorización de los gráficos de control en el estudio de simulación.

Como se ha visto en el capítulo 2, para diseñar la fase de monitorización de una carta de control es necesario establecer los límites de control, para determinar el percentil de la distribución del estadístico ( $T^2$  Hotelling,  $L^1$  y  $L^2$  estandarizadas). En este capítulo se presentarán las técnicas de remuestreo bootstrap, iniciando por el remuestreo clásico en el campo multivariante y extendiéndonos al bootstrap suavizado que es una alternativa para los datos funcionales como lo indica en [28]. También se considerará la técnica de permutaciones que se presenta en [29], [30], [31]. Finalmente, se presentarán las medidas de profundidad que permitirán en una primera instancia recortar un determinado porcentaje de las curvas atípicas como se muestra en [32].

### 3.1 Espacios de Hilbert separables

Sea  $\mathbb{H}$  un espacio separable<sup>1</sup> de Hilbert. A continuación introduciremos la notación de las operaciones de suma, producto con un escalar y el producto inercial en  $\mathbb{H}$ .

- La suma de dos elementos  $f, g \in \mathbb{H}$  se denota por  $f \oplus g$
- El producto escalar de un elemento  $f \in \mathbb{H}$  por un escalar  $\lambda \in \mathbb{R}$  es denotado por  $\lambda \odot f$
- La resta entre dos elementos  $f, g \in \mathbb{H}$  está totalmente definido por las operaciones de suma y del producto escalar. Para introducir esta operación en  $\mathbb{H}$ , denotamos así,  $\ominus := \oplus(-1) \odot$
- El producto inercial entre dos elementos  $f, g \in \mathbb{H}$  está denotado por  $\langle \cdot, \cdot \rangle_{\mathbb{H}}$

**Definición 3.1.1.** Una  $\mathbb{H}$ - variable aleatoria en un espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$  una función  $\mathcal{X} : \Omega \rightarrow \mathbb{H}$  tal que  $\langle \mathcal{X}, f \rangle_{\mathbb{H}}$  es medible para todo  $f \in \mathbb{H}$ .

**Definición 3.1.2.** Sea  $\mathcal{X}$  una  $\mathbb{H}$  variable aleatoria en el espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$ . La integral de Bochner de  $\mathcal{X}$  con respecto a la medida de probabilidad  $\mathbb{P}$  es el único elemento  $\mathbb{E}_{\mathbb{H}}[\mathcal{X}]$  de  $\mathbb{H}$  tal que

<sup>1</sup>véase el capítulo B, teorema B.0.6



$$\langle \mathbb{E}[\mathcal{X}], f \rangle_{\mathbb{H}} = \mathbb{E}_{\mathbb{R}}[\langle \mathcal{X}, f \rangle_{\mathbb{H}}]$$

para todo  $f \in \mathbb{H}$  con  $\mathbb{E}_{\mathbb{R}}$  es el operador de esperanza para variables aleatorias en  $\mathbb{R}$ .

**Definición 3.1.3.** Sean  $f, g$  dos elementos de un espacio separable de Hilbert  $\mathbb{H}$ . El operador del producto tensorial  $(f \otimes g)_{\mathbb{H}} : \mathbb{H} \rightarrow \mathbb{H}$  está definida como sigue

$$\begin{aligned} (f \otimes g)_{\mathbb{H}} : \mathbb{H} &\longrightarrow \mathbb{H} \\ h &\mapsto \langle f, h \rangle_{\mathbb{H}} \odot g \end{aligned}$$

## 3.2 Integral vectorial

Supongamos ahora que tenemos una función  $f$  en un espacio de medida  $(E, B, \mu)$  que toma valores en un espacio de Banach  $\mathbb{X}$ . En tal caso, cuando  $\mathbb{X} = \mathbb{R}$  estamos familiarizados con la integral de Lebesgue. Se requiere extender esta idea a un espacio de Banach general. El objetivo de esta sección es exponer en forma concisa la teoría sobre la integral vectorial que construyó Salomon Bochner en su paper *Integration von Functionen* [33] en 1933. En particular esta integral será relevante en la definición del elemento medio y el operador de covarianza de los elementos aleatorios definidos en espacios de Hilbert.

### 3.2.1 Integral de Bochner

**Definición 3.2.1.** Una función  $f : E \mapsto \mathbb{X}$  se llama función simple si se puede representar como sigue

$$f(\omega) = \sum_{i=1}^k I_{E_i}(\omega) g_i \tag{3.1}$$

para un  $k$  finito,  $E_i \in \mathcal{B}$  y  $g_i \in \mathbb{X}$ .

**Definición 3.2.2.** Sea una función simple  $f(\omega) = \sum_{i=1}^k I_{E_i}(\omega)g_i$  con  $\mu(E_i) < \infty$  para todo  $i$  se dice que es integrable y la integral de Bochner se define como sigue

$$\int_E f d\mu = \sum_{i=1}^k \mu(E_i)g_i \quad (3.2)$$

La definición anterior se extiende a una función general medible de  $E$  a  $\mathbb{X}$  como sigue.

**Definición 3.2.3.** Una función medible  $f$  es llamada una integral de Bochner si existe una sucesión  $\{f_n\}$  de funciones simples bochner integrables tal que

$$\lim_{n \rightarrow \infty} \int_E \|f_n - f\| d\mu = 0 \quad (3.3)$$

En este caso, la integral de Bochner de  $f$  está definida así

$$\int_E f d\mu = \lim_{n \rightarrow \infty} \int_E f_n d\mu \quad (3.4)$$

**Teorema 3.2.1.** Si  $f$  es una función medible de  $E$  en  $\mathbb{X}$  con

$$\int_E \|f\| d\mu < \infty \quad (3.5)$$

Supongamos que para cada  $n$  existe un subespacio de dimensión finita  $\mathbb{X}_n$  de  $\mathbb{X}$  tal que

$$\lim_{n \rightarrow \infty} \int \|f - g_n\| d\mu = 0 \quad (3.6)$$

para cualquier  $g_n$  función medible del subespacio  $\mathbb{X}_n$ . Entonces existen funciones simples Bochner integrables  $f_n$  tal que se satisface 3.3.

*Demostración.* La demostración véase en [34, p. 43].

El siguiente resultado muestra que la existencia de  $g_n$  en 3.6 está garantizada si  $\mathbb{X}$  es un espacio de Hilbert separable.

**Teorema 3.2.2.** *Supongamos que  $\mathbb{H}$  es un espacio de Hilbert separable y  $f$  es una función medible de  $E$  en  $\mathbb{H}$  con  $\int_E \|f\| d\mu < \infty$ . Entonces,  $f$  es Bochner integrable.*

*Demostración.* La demostración véase en [34, p. 44].

### 3.3 Elementos aleatorios en espacios de Hilbert

#### 3.3.1 Medidas de probabilidad en un espacio de Hilbert

En esta sección trabajaremos en el contexto del espacio de Hilbert separable  $\mathbb{H}$  con norma y producto inercial asociados  $\|\cdot\|$ ,  $\langle \cdot, \cdot \rangle$ . Topológicamente, se puede ver a  $\mathbb{H}$  como un espacio métrico, donde

$$d(f, g) = \|f - g\| = \langle f - g, f - g \rangle^{\frac{1}{2}}$$

La  $\sigma$ -álgebra boreliana de cualquier espacio topológico  $\mathbb{X}$  es la más pequeña  $\sigma$ -álgebra que contiene todos los conjuntos abiertos (respecto a la norma basada en la métrica) de  $\mathbb{X}$  y los notamos por  $\mathcal{B}(\mathbb{X})$ . La más pequeña  $\sigma$ -álgebra que contiene a la clase  $\mathcal{C}$  de conjuntos se la nota también  $\sigma(\mathcal{C})$ . Sea  $\mathcal{M}$  la clase de todos los conjuntos de la forma  $\{x \in \mathbb{H} : \langle x, f \rangle \in B\}$  para todo  $f \in \mathbb{H}$  y  $B$  un conjunto abierto de  $\mathbb{R}$ .

**Teorema 3.3.1.** *La  $\sigma$ -álgebra  $\sigma(\mathcal{M})$  y  $\mathcal{B}(\mathbb{H})$  son iguales.*

*Demostración.* La demostración véase en [34, p. 176].

El siguiente resultado proporciona una caracterización específica del espacio de Hilbert de medibilidad que necesitaremos en secciones posteriores de este capítulo.

**Teorema 3.3.2.** *Sea  $\mathcal{X}$  una función sobre el espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$  a  $(\mathbb{H}, \mathcal{B}(\mathbb{H}))$ . Entonces,  $\mathcal{X}$  es medible si  $\langle \cdot, \cdot \rangle$  es medible para todo  $f \in \mathbb{H}$*

*Demostración.* La demostración véase en [34, p. 177].

En lo posterior nos referiremos a funciones medibles que toman valores en un espacio de Hilbert  $\mathbb{H}$  como elementos aleatorios de  $\mathbb{H}$ .

### 3.3.2 Media y covarianza de un elemento aleatorio en un espacio de Hilbert

Sea  $\mathcal{X}$  un elemento aleatorio en un espacio de Hilbert separable definido en el espacio de probabilidad  $(\Omega, \mathcal{F}, \mathbb{P})$ . El teorema 3.2.2 sugiere una noción para la media de un elemento aleatorio.

**Definición 3.3.1.** *Si  $\mathbb{E}(\|\mathcal{X}\|) < \infty$ , la media de un elemento aleatorio  $\mathcal{X}$ , está definido por la integral de Bochner.*

$$m = \mathbb{E}(\mathcal{X}) := \int_{\Omega} \mathcal{X} d\mathbb{P}$$

La definición 3.3.1 provee de una extensión natural de la media de una variable aleatoria al caso de elementos aleatorios. En términos generales, una “suma ponderada” de las posibles realizaciones de  $\mathcal{X}$  devuelve otro elemento no aleatorio de  $\mathbb{H}$ .

Asumiendo que la esperanza existe, tomaremos  $\mathbb{E} \|\mathcal{X} - m\|^2$  como la varianza de un elemento aleatorio. En este sentido tenemos el siguiente resultado.

**Teorema 3.3.3.** *Supongamos que  $\mathbb{E} \|\mathcal{X}\|^2 < \infty$ . Entonces,*

$$\mathbb{E} \|\mathcal{X} - m\|^2 = \mathbb{E} \|\mathcal{X}\|^2 - \|m\|^2$$

donde  $m$  es la media de  $\mathcal{X}$

*Demostración.* La demostración véase en [34, p. 179].

El siguiente paso es desarrollar un concepto de covarianza para  $\mathcal{X}$ . En este sentido, recordando que la covarianza para un  $p$ -vector aleatorio es  $X$

$$\mathbb{E}[(X - \mathbb{E}X)(X - \mathbb{E}X)^T] = \mathbb{E}[(X - \mathbb{E}X) \otimes (X - \mathbb{E}X)] \quad (3.7)$$

Basados en la idea anterior, formularemos el concepto de covarianza en espacios de Hilbert. En específico, si  $\mathcal{X}$  es un elemento aleatorio de un espacio de Hilbert  $\mathbb{H}$ , definimos el operador de covarianza <sup>2</sup> correspondiente como sigue.

**Definición 3.3.2.** *Supongamos que  $\mathbb{E} \|\mathcal{X}\|^2 < \infty$ . Entonces, el operador de covarianza para  $\mathcal{X}$  es un elemento de  $\mathfrak{B}_{HS}(\mathbb{H})$  dado por la integral de Bochner.*

$$\mathcal{K} = \mathbb{E}[(\mathcal{X} - m) \otimes (\mathcal{X} - m)] := \int_{\Omega} (\mathcal{X} - m) \otimes (\mathcal{X} - m) d\mathbb{P}$$

Si  $\mathcal{X}(\omega)$  está en  $\mathbb{H}$ , entonces  $(\mathcal{X}(\omega) - m) \otimes (\mathcal{X}(\omega) - m)$  es un operador de Hilbert Schmidt con norma  $\|\mathcal{X}(\omega) - m\|^2$ . Así,  $(\mathcal{X} - m) \otimes (\mathcal{X} - m)$  es un elemento aleatorio del espacio de Hilbert  $\mathfrak{B}_{HS}(\mathbb{H})$ . Por el teorema 3.3.3, la suposición  $\mathbb{E} \|\mathcal{X}\|^2 < \infty$  implica que la esperanza de la norma HS de  $(\mathcal{X} - m) \otimes (\mathcal{X} - m)$  es finita. Además, como  $\mathfrak{B}_{HS}(\mathbb{H})$  es un espacio de Hilbert Separable, se sigue que por el teorema 3.2.2,  $\mathcal{K}$  está bien definido por la integral de Bochner que da como resultado un elemento de  $\mathfrak{B}_{HS}(\mathbb{H})$ .

---

<sup>2</sup>Véase en la sección B

### 3.4 Estadístico basado en la distancia de Mahalanobis: $T^2$ de Hotelling

La aplicación de las distancias en estadística están latentes en algunas aplicaciones, (como por ejemplo, en el test ji-cuadrado de K.Pearson o en el test  $t$  de Student) las discrepancias entre lo observado y esperado se mide mediante un estadístico que en el fondo es una distancia. La distancia debida a Mahalanobis interviene en la propia definición de la distribución normal multivariante, análisis discriminante, en la detección de datos atípicos, y particularmente, en el estadístico  $T^2$  de Hotelling.

El estadístico  $T^2$  de Hotelling surge en la estadística multivariante al realizar pruebas de las diferencias de medias (multivariadas) de diferentes poblaciones, donde las pruebas para problemas univariados hacen uso de la prueba  $t$ . La distribución lleva el nombre de Harold Hotelling, quien lo desarrolló como una generalización de la distribución  $t$  de Student. A continuación se definirá dicho estadístico extendido al contexto de los datos funcionales pasando por el caso multivariante.

El estadístico  $T^2$  multivariante de Hotelling está estrechamente relacionado con el estadístico  $t$  de Student univariante. En efecto, para cualquier dirección  $a \in \mathbb{R}^p \setminus \{0\}$  este estadístico se calcula como sigue

$$t_a = \sqrt{n} \frac{a^T(m_n - m)}{\sqrt{a^T \Sigma_n a}} \quad (3.8)$$

La definición alternativa del estadístico  $T^2$  de Hotelling es

$$T^2 := \max_{a \in \mathbb{R}^p} t_a^2 \quad (3.9)$$

Cuando  $n > p$  la matriz de covarianza muestral  $\Sigma_n$  es de rango completo. Además  $Im(\Sigma_n) = \mathbb{R}$ . Por tanto independientemente del tamaño de muestra  $n$ , el estadístico

$T^2$  de Hotelling se puede definir en su forma más general como se define a continuación

**Definición 3.4.1.** Sean  $X_1, \dots, X_n \in \mathbb{R}^p$  variables aleatorias i.i.d con media  $m$  y matriz de covarianza  $\Sigma$ . El estadístico  $T^2$  de Hotelling se define como sigue:

$$T^2 = n \max_{a \in \text{Im}(\Sigma_n) - \{0\}} \frac{\langle a, \mathcal{D}_n a \rangle_{\mathbb{R}^p}}{\langle a, \Sigma_n a \rangle_{\mathbb{R}^p}} \quad (3.10)$$

Donde  $\mathcal{D}_n = (m_n - m)(m_n - m)^T$ ,  $\Sigma_n$  es la matriz de covarianza y es el producto innercial en  $\mathbb{R}^p$

En particular cuando  $p \geq n$  se ha probado en [35] que

$$T^2 = n(m_n - m)^T \Sigma_n^+ (m_n - m) \quad (3.11)$$

Donde  $\Sigma_n^+$  es la inversa Moore-Penrosed de la matriz de covarianza muestral  $\Sigma_n$ .

### 3.4.1 Estadístico $T^2$ en el espacio de Hilbert separable

**Definición 3.4.2.** Sean  $\mathcal{X}_1, \dots, \mathcal{X}_n$  una muestra de  $n$  i.i.d  $\mathbb{H}$ -variables aleatorias en  $(\Omega, \mathcal{F}, \mathbb{P})$ , tal que  $\mathbb{E}[\|\mathcal{X}_i\|_{\mathbb{H}}^2] < +\infty$  para todo  $i = 1, \dots, n$ . Sea la media  $m$  y el operador de  $\mathcal{K}$  de  $\mathcal{X}_i$ , se define:

La **media muestral**  $m_n$

$$m_n = \frac{1}{n} \odot \bigoplus_{i=1}^n \mathcal{X}_i$$

El **operador muestral**  $\mathcal{K}_n$  se define como

$$\mathcal{K}_n = \frac{1}{n-1} \odot \bigoplus_{i=1}^n (\mathcal{X}_i \ominus m_n) \otimes_{\mathbb{H}} (\mathcal{X}_i \ominus m_n)$$

$m_n$  y  $\mathcal{K}_n$  son respectivamente  $\mathbb{H}$ -variable aleatoria y  $\mathfrak{B}_{HS}(\mathbb{H})$  variable aleatoria.

**Definición 3.4.3.** Usando las notaciones de la definición anterior, el operador muestral de pérdida de error cuadrático medio  $\mathcal{D}_n$  se define así:

$$\mathcal{D}_n = (m_n \ominus m) \otimes_{\mathbb{H}} (m_n \ominus m)$$

**Definición 3.4.4.** Sean  $\mathcal{X}_1, \dots, \mathcal{X}_n$  i.i.d  $\mathbb{H}$  variables aleatorias con media  $m$  y operador de covarianza  $\mathcal{K}$ . El operador  $T^2$  de Hotelling se define como sigue:

$$T^2 = n \max_{f \in \text{Im}(\mathcal{K}_n) \setminus \{0\}} \frac{\langle f, \mathcal{D}_n f \rangle}{\langle f, \mathcal{K}_n f \rangle} \quad (3.12)$$

lo que es equivalente a:

$$T^2 = n \langle m_n \ominus m, \mathcal{K}_n^+(m_n \ominus m) \rangle_{\mathbb{H}}$$

donde  $\mathcal{K}_n$  es el operador de covarianza y  $\mathcal{K}_n^+$  es el operador inverso generalizado  $\mathcal{K}_n$  y  $\mathcal{D}_n$  es el operador muestral de pérdida de error medio cuadrático.

### 3.4.2 Cálculo numérico del estadístico $T^2$ de Hotelling

En la definición 3.4.4 se garantiza la existencia del estadístico  $T^2$  de Hotelling en espacios de Hilbert separables, sin embargo, es importante ilustrar su aplicación en espacios de Hilbert específicos usados en casos prácticos. La idea central es tener una sucesión del estadístico  $T^2$  finito dimensional que converge casi seguramente al estadístico  $T^2$  de Hotelling.

Sea  $\mathbb{H}$  el espacio separable de Hilbert, entonces existe una sucesión de subespacios  $\{V_p\}_{p \geq 1}$  tal que  $V_p \subset \mathbb{H}$ ,  $\dim(V_p) = p$  y,

$$\lim_{p \rightarrow \infty} \inf_{w_p \in V_p} \|h \ominus w_p\|_{\mathbb{H}} = 0 \quad \forall h \in \mathbb{H}$$



La última propiedad nos quiere decir que, a medida que la dimensionalidad de  $p$  del espacio  $V_p$  tiende al infinito, dicho espacio tiende a cubrir  $\mathbb{H}$ . Dada la existencia de una sucesión de subespacios entonces es posible definir la sucesión  $\{T_p\}_{p \geq 1}$  como sigue

$$T_p^2 = n \max_{f \in \text{Im}(\mathcal{K}_n) \cap V_p \setminus \{0\}} \frac{\langle f, \mathcal{D}_n f \rangle_{\mathbb{H}}}{\langle f, \mathcal{K}_n f \rangle_{\mathbb{H}}} \quad (3.13)$$

**Teorema 3.4.1.** *Un espacio separable de Hilbert y  $\{V_p\}_{p \geq 1}$  una sucesión de subespacios tal que  $V_p \subset \mathbb{H}$ ,  $\dim(V_p) = p$  y  $\lim_{p \rightarrow \infty} \inf_{w_p \in V_p} \|h \ominus w_p\|_{\mathbb{H}} = 0$  para todo  $h \in \mathbb{H}$ . Sean  $\mathcal{X}_1, \dots, \mathcal{X}_n$  una muestra de  $n$  i.i.d.  $\mathbb{H}$ -variables aleatorias con media  $\mu$  y operador de covarianza  $\mathcal{K}$  tal que  $\mathbb{E}_{\mathbb{R}}[\|\mathcal{X}_i\|_{\mathbb{H}}^2] < \infty$ . Entonces, la sucesión  $\{T_p^2\}_{p \geq 1}$  definida en la ecuación 3.13 tiene las siguientes propiedades*

a) Si  $(e_1, \dots, e_p)$  es una base de  $V_p$ , sea  $\mathbf{W}$  una matriz simétrica invertible  $p \times p$  tal que  $\mathbf{W}_{\mathbf{jk}} := \langle \mathbf{e}_j, \mathbf{e}_k \rangle_{\mathbb{H}}$ . Así

$$\begin{aligned} \mathcal{X}_i &:= \mathbf{W}^{-1}(\langle \mathcal{X}_i, \mathbf{e}_1 \rangle_{\mathbb{H}}, \dots, \langle \mathcal{X}_i, \mathbf{e}_p \rangle_{\mathbb{H}})' \\ \mathbf{m}_n &:= \mathbf{W}^{-1}(\langle \mathbf{m}_n, \mathbf{e}_1 \rangle_{\mathbb{H}}, \dots, \langle \mathbf{m}_n, \mathbf{e}_p \rangle_{\mathbb{H}})' \\ \mathbf{m} &:= \mathbf{W}^{-1}(\langle \mathbf{m}, \mathbf{e}_1 \rangle_{\mathbb{H}}, \dots, \langle \mathbf{m}, \mathbf{e}_p \rangle_{\mathbb{H}})' \\ \Sigma_n &:= \frac{1}{n-1} \sum_{i=1}^n (\mathcal{X}_i - \mathbf{m}_n)(\mathcal{X}_i - \mathbf{m}_n)' \end{aligned}$$

Entonces,

$$T_p^2 = n(\mathbf{m}_n - \mu)' \mathbf{W}^{\frac{1}{2}} (\mathbf{W}^{\frac{1}{2}} \Sigma_n \mathbf{W}^{\frac{1}{2}})^+ \mathbf{W}^{\frac{1}{2}} (\mathbf{m}_n - \mu)$$

b)  $T_p^2 \xrightarrow{a.s.} T^2$ , cuando  $p \rightarrow \infty$ , donde  $T^2$  es el estadístico  $T^2$  de Hotelling definido en el espacio original de la muestra de  $\mathcal{X}_1, \dots, \mathcal{X}_n$  i.i.d.  $\mathbb{H}$  variables aleatorias

*Demostración.* La demostración véase en [27, p. 29].

### 3.4.3 Estadístico basado en la distancia de Mahalanobis a dos muestras

El estadístico basado en la distancia de Mahalanobis a dos muestras está definido como sigue

$$T_0^2 = \left( \frac{1}{n_1} + \frac{1}{n_2} \right)^{-1} \max_{f \in \text{Im}(\mathcal{K}_{n_{pooled}}) \setminus \{0\}} \frac{\langle f, \mathcal{D}_{n_0} f \rangle}{\langle f, \mathcal{K}_{n_{pooled}} f \rangle} \quad (3.14)$$

Donde  $m_{n_1}$  y  $m_{n_2}$  son las medias de las muestras,  $\mathcal{D}_{n_0}$  es el operador de pérdida de error medio cuadrático, bajo la hipótesis nula, i.e.  $\mathcal{D}_{n_0} = (m_{n_1} \ominus m_{n_2}) \otimes (m_{n_1} \ominus m_{n_2})$  y  $\mathcal{K}_{n_{pooled}}$  el operador del muestra de covarianza definido como sigue

$$\begin{aligned} \mathcal{K}_{n_{pooled}} : \mathbb{H} &\longmapsto \mathbb{H} \\ f &\longmapsto \frac{n_1 - 1}{n_1 + n_2 - 2} \odot (\mathcal{K}_{n_1} f) \oplus \frac{n_2 - 1}{n_1 + n_2 - 2} \odot (\mathcal{K}_{n_2} f), \end{aligned}$$

donde  $\mathcal{K}_{n_1}$  y  $\mathcal{K}_{n_2}$  son los operadores de covarianza de las muestras. Alcanza el máximo para  $f = \mathcal{K}_{n_{pooled}}^+(m_{n_1} \ominus m_{n_2})$  como se muestra en [27].

### 3.4.4 Distancias para datos funcionales

Las distancias más usadas en  $\mathbb{R}^p$  es la familia de las distancias de Minkowski, en particular  $L^1$  y  $L^2$ ; naturalmente, se han extendido al contexto funcional y se las ha usado ampliamente en las aplicaciones del análisis de datos funcionales como se ve en los trabajos de Ramsay & Silverman [36], Horvath y Kokoszka [37].

El espacio de Hilbert separable más utilizado para el análisis de datos funcionales es el espacio de Sobolev, espacio de (clases de equivalencia) funciones cuadrado integrables sobre  $T$ , tradicionalmente conocido como el espacio  $L^2(T)$ . Dada la popularidad de este espacio en el análisis de datos funcionales, algunos estadísticos lo surgieron en la literatura para hacer inferencias sobre la función media de una población de curvas o sobre la diferencia entre las medias funcionales de dos poblaciones de curvas, cuando son curvas aleatorias en el espacio  $L^2(T)$ .

A continuación se presentan brevemente las definiciones de las distancias estandarizadas  $L^1$  y  $L^2$  presentados por Hall y Tajvidi [4], con la finalidad de diseñar la fase de monitorización a partir de la comparación de la media de la muestra bajo control y la muestra de monitoreo, considerando las distancias mencionadas anteriormente y la distancia de Mahalanobis extendida al contexto funcional.

$$L_s^1 = \int_T \frac{|m_n - m(t)|}{\sigma_n^2(t)} dt \quad (3.15)$$

$$L_s^2 = \int_T \frac{(m_n - m(t))^2}{\sigma_n^2(t)} dt \quad (3.16)$$

donde se define  $\sigma_n^2 : T \mapsto \mathbb{R}$  como la función de varianza de la muestra puntual siendo  $(n-1)\sigma_n^2(t) = \sum_{i=1}^n (\mathcal{X}_i(t) - m(t))^2$

### 3.5 Técnicas de remuestreo bootstrap

En el año 1979, Bradley Efron desarrolla la técnica bootstrap. Concretamente, la idea básica sugerida por Efron en [38] es la siguiente. Si una muestra aleatoria contiene la máxima información disponible sobre la población, ¿por qué no proceder como si la muestra fuese la población y, entonces, estimar la distribución muestral de un estadístico generando nuevas muestras mediante un muestreo aleatorio con reposición, a partir de los datos de la muestra original?. El remuestreo bootstrap utiliza la

muestra como mecanismo generador de nuevas muestras y permite ampliar el uso de las simulaciones de Monte Carlo, al no precisar el conocimiento de los parámetros poblacionales, ni realizar supuestos sobre estos [39].

Inicialmente Efron describe el bootstrap como una técnica útil para aproximar la distribución de un estadístico, la principal ventaja es que no se requiere hipótesis sobre el mecanismo generador de los datos, además la implementación computacional suele ser sencilla en comparación con otros métodos. En las siguientes secciones se exponen los esquemas de bootstrap clásico y suavizado.

### 3.5.0.1 Bootstrap clásico

Supongamos que tenemos una muestra aleatoria  $X = (X_1, \dots, X_n)$  de una función de distribución  $F$  desconocida y  $T_n(X, F)$  un estadístico. Sea  $F_n$  la distribución empírica construida a partir de la muestra.

El bootstrap uniforme es aquel en el que se reemplaza la distribución poblacional (desconocida) por la distribución empírica

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}\{X_i \leq x\}$$

Es decir,  $\hat{F} = F_n$ , y, por lo tanto  $R^* = R(\mathbf{X}^*, \mathbf{F}_n)$

Así pues, en este caso el algoritmo bootstrap uniforme o clásico es el siguiente

1. Para  $i \in \{1, \dots, n\}$  arrojar  $X_i^*$  a partir de  $F_n$ , es decir,  $P^*(X_i^* = X_j) = \frac{1}{n}$  para  $j \in \{1, \dots, n\}$
2. Obtener  $\mathbf{X}^* = (\mathbf{X}_1^*, \dots, \mathbf{X}_n^*)$
3. Calcular  $R^* = R(\mathbf{X}^*, \mathbf{F}_n)$

### 3.5.1 Bootstrap suavizado

Cuando la distribución poblacional  $F$  es continua es lógico incorporar dicha información al bootstrap. Eso significa que la función de distribución tiene una función de densidad asociada, relacionadas mediante la expresión:  $f(x) = F'(x)$ . Para ello, debemos utilizar un estimador de la función de densidad y remuestrear de él.

Pasamos a considerar brevemente el problema de estimar, no paramétricamente, la función de densidad,  $f$ , de una población a partir de la muestra  $(X_1, X_2, \dots, X_n)$ , procedente de la misma. En ese contexto es recomendable utilizar el estimador tipo núcleo propuesto por Parzen [40] y Rosenblatt [41], que viene dado por

$$\hat{f}_h(x) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i)$$

donde

$$K_h(u) = \frac{1}{h} K\left(\frac{u}{h}\right)$$

$K$  es una función núcleo (usualmente una densidad simétrica en torno a cero) y  $h \geq 0$  es un parámetro suavizado, llamado ventana, que regula el tamaño del entorno que se usa para llevar a cabo la estimación. Así, eligiendo como función  $K$  la densidad de una  $\mathcal{U}(-1, 1)$ , el estimador de Parzen- Rosenblatt resulta

$$\begin{aligned} \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} \mathbb{1}\left\{\frac{x - X_i}{h} \in (-1, 1)\right\} &= \frac{1}{2nh} \sum_{i=1}^n \mathbb{1}\{X_i \in (x - h, x + h)\} \\ &= \frac{\#\{X_i \in (x - h, x + h)\}}{2nh} \end{aligned} \quad (3.17)$$

que no es más que la frecuencia relativa de datos  $\mathbf{X}_i$  en el intervalo  $(x - h, x + h)$  dividida entre la longitud del intervalo en cuestión ( $2h$ ).

Es habitual exigir que la función núcleo  $K$  sea una función no negativa, su integral uno:

$$K(u) \geq 0, \quad \forall u, \quad \int_{-\infty}^{\infty} K(u)du = 1$$

Además también es frecuente exigir que  $K$  sea una función simétrica ( $K(-u) = K(u)$ )

La elección de la función  $K$  no tiene gran impacto en las propiedades del estimador (salvo sus condiciones de regularidad: continuidad, diferenciabilidad, etc) la elección del parámetro de suavizado sí es muy importante para una correcta estimación. En otras palabras, el tamaño del entorno usado para la estimación no paramétrica debe ser adecuado.

El método bootstrap suavizado procede de la siguiente forma:

1. A partir de la muestra  $(X_1, X_2, \dots, X_n)$  y utilizando un valor  $h > 0$  como parámetro de suavizado, se calcula el estimador de Parzen- Rosenblatt  $\hat{f}_h$
2. Calcular  $R^* = R(\mathbf{X}^*, \hat{\mathbf{F}}_h)$
3. Repetir B veces los paso 2-3 para obtener las réplicas bootstrap  $R^{*(1)}, \dots, R^{*(B)}$

Para llevar a cabo un bootstrap que remuestree a partir del estimador  $\hat{f}_h(x)$  es útil pensar en dicho estimador como una combinación lineal convexa de funciones de densidad  $K_h(x - X_i)$ , cada una con coeficiente  $\frac{1}{n}$  en dicha combinación lineal. Gracias a esa representación podemos simular valores  $X^*$ , procedentes de  $\hat{f}_h(x)$ .

En un primer paso elegiremos (aleatoriamente y con equiprobabilidad) cuál de los índices  $i \in \{1, \dots, n\}$  vamos a considerar y en segundo paso simularemos  $X^*$  a partir de la densidad  $K_h(\cdot - X_i)$ . Esta última fase puede relacionarse fácilmente con la simulación de un valor,  $V$ , con densidad  $K$ , sin más que hacer  $X_i + hV$ . Así el paso 2 del algoritmo previo puede llevarse a cabo mediante el siguiente procedimiento:

2. Para cada  $i = 1, \dots, n$  arrojar  $U_i \sim \mathbf{U}(0, 1)$  y  $V_i$  con densidad  $K$  y luego hacer  $X_i^* = X_{\lfloor nU_i \rfloor + 1} + hV_i$ .

### 3.5.2 Permutaciones

En un principio esta metodología fue estudiada por diferentes autores [29],[30],[31] y frecuentemente usada para el caso de comparación de medias. Este método es similar a la técnica bootstrap aunque la diferencia fundamental es que para bootstrap el remuestreo es con reemplazo mientras que para el caso de permutaciones no.

Si notamos  $X_1, X_2, \dots, X_n$  y  $Y_1, Y_2, \dots, Y_m$  dos muestras independientes tales que  $X \sim F$  y  $Y \sim G$ . La prueba se establece de la siguiente manera:

$H_0$ : X y Y tienen características similares

$H_1$ :  $H_0$  no es cierta

De acuerdo a [42] el procedimiento de la prueba se puede describir como sigue a continuación:

Sea  $D = D(X_1, \dots, X_n, Y_1, \dots, Y_m)$  una prueba estadística adecuada para diferenciar  $H_1$  y  $H_0$ .

Sea  $d$ , el valor observado cuando el estadístico  $D$  se aplica a los datos originales.

Sea  $N = n + m$ , y considerando todas las permutaciones posibles  $N!$  de la muestra conjunta  $X_1, \dots, X_n, Y_1, \dots, Y_m$ . Para cada permutación, se calcula el estadístico de prueba  $D$  y se denota estos valores como  $D_1, D_2, \dots, D_{N!}$

Bajo la hipótesis nula cada uno de los valores obtenidos en el paso anterior son igualmente probables. La distribución que coloca el peso  $\frac{1}{N!}$  en cada uno de los valores  $D_1, \dots, D_{N!}$  se conoce como la distribución de permutación de  $D$ . Suponiendo que la

hipótesis nula se rechaza para valores grandes de  $D$ , entonces el p-valor de la prueba es

$$p_{valor} = \frac{1}{N!} \sum_{j=1}^{N!} \mathbb{1}(D_j > d)$$

Supongamos que se requiere realizar una prueba basada en permutación y calcular el valor de  $p$  asociado, bajo la hipótesis nula de que las distribuciones evaluadas  $X$  y  $Y$  son las mismas.

Para ello, los elementos  $X$  y  $Y$  son unidos y permutados, después de la permutación los elementos se dividen en dos grupos  $X$  y  $Y$ , conservando el tamaño original de las muestras. La distribución del estadístico estudiado bajo  $H_0$  se obtiene calculando el valor de dicho estadístico para todas las posibles reorganizaciones de las observaciones en los distintos grupos.

### 3.6 Medidas de Profundidad

Las medidas de profundidad se introdujeron originalmente en el análisis multivariante para medir la centralidad de un punto  $X \in \mathbb{R}^d$  con respecto a una nube de datos generada a partir de una distribución de probabilidad  $F \in \mathbb{R}^d$ .

La profundidad funcional permite ordenar una muestra de curvas desde el centro hacia afuera. La curva con mayor profundidad se corresponde con la mediana mientras que los atípicos se definen como las curvas de menor profundidad.

Las profundidades proporcionan una manera de ordenar puntos en el espacio Euclídeo desde el centro hacia el exterior, de tal manera que los punto más cercanos al centro tienen una mayor profundidad que los puntos localizados en los extremos que tienen menor profundidad.



La profundidad de Fraiman y Muniz [43], es una de las primeras medidas de profundidad introducidas para datos funcionales.

**Definición 3.6.1.** Sean  $\mathcal{X}_1, \dots, \mathcal{X}_n$  i.i.d, realizaciones de la variable aleatoria funcional  $\mathcal{X}(t)$  con dominio  $T = [a, b]$  y sea  $D$  una medida de profundidad en  $\mathbb{R}$ . Para cada  $t_0 \in T$ , se considera  $z_i(t_0) = D(\mathcal{X}_i(t_0))$  la profundidad univariante del dato  $i$  en  $t_0$  con respecto a  $\mathcal{X}_1(t_0), \dots, \mathcal{X}_n(t_0)$ . Entonces se define la profundidad de FM para el  $i$ -ésimo dato como:

$$FM_i = FMD(\mathcal{X}_i) = \int_a^b z_i(t) dt$$

La medida de profundidad univariante definida por Fraiman y Muniz en [43] es la siguiente:

$$1 - \left| \frac{1}{2} - F_{n,t}(\mathcal{X}_i(t)) \right|$$

donde

$$F_{n,t} = \frac{1}{n} \sum_{k=1}^n \mathbb{1}\{\mathcal{X}_k(t) \leq \mathcal{X}_i(t)\}$$

Entonces, la profundidad funcional de Fraiman y Muniz de la curva  $\mathcal{X}$  con respecto al conjunto  $\mathcal{X}_1, \dots, \mathcal{X}_n$  está dada por:

$$FMD(\mathcal{X}_i) = \int_a^b z_i^{FM1}(t) dt$$

**Definición 3.6.2.** Sean  $\mathcal{X}_1, \dots, \mathcal{X}_n$  i.i.d realizaciones de la variable aleatoria  $\mathcal{X}$  y  $K : \mathbb{R}^+ \rightarrow \mathbb{R}^+$  una función kernel y  $h$  un parámetro ventana. Entonces la profundidad funcional modal MD para el  $i$ -ésimo dato, definido por [44]

$$MD_i = MD(\mathcal{X}_i) = \sum_{j=1}^n K \left( \frac{\|\mathcal{X}_i - \mathcal{X}_j\|}{h} \right)$$

La profundidad modal es una medida de cuántos datos hay en un vecindario de la curva y se asemeja el estimador densidad del kernel no paramétrico pero con la principal diferencia de que el parámetro ventana no necesariamente tiende a cero a medida que  $n$  aumenta. El objetivo de esta medida de profundidad es proporcionar

rangos que no cambien para anchos de banda lo suficientemente grandes.

Desde el punto de vista práctico, hay que elegir una norma funcional, una función núcleo y un  $h$ . En [44], se recomienda utilizar la norma  $L^2$ . Además, se recomienda usar la función núcleo

$$K(t) = \frac{2}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right), t > 0$$

El parámetro  $h$  recomendado es el décimo quinto percentil de la distribución empírica de  $\|\mathcal{X}_i - \mathcal{X}_j\|$  con  $i, j = 1, \dots, n$ . Dado que el interés está en los valores alrededor del centro de la distribución, que no son muy sensibles a la elección del  $h$ , en [45] se recomienda que  $h$  no sea muy pequeño.

Se cuenta con la extensión de la medida de profundidad modal funcional para el caso p-variado funcional  $MD^p$ . En [46] consideran una medida de profundidad de los datos funcionales bajo proyecciones. La idea básica es proyectar cada curva funcional a lo largo de una dirección aleatoria.

**Definición 3.6.3.** Sean  $\mathcal{X}_1, \dots, \mathcal{X}_n$  i.i.d realizaciones de la variable aleatoria funcional  $\mathcal{X}(t)$  con dominio  $T = [a, b]$ ,  $D$  una medida de profundidad univariante y sea  $h$  una realización aleatoria que pertenece a un proceso independiente  $\mathcal{H}$ , y  $P_i^h = \langle h, \mathcal{X}_i \rangle \in \mathbb{R}$  es la proyección de  $\mathcal{X}_i$  a lo largo de la dirección  $h$ , esto es:

$$P_i^h = \langle h, \mathcal{X}_i \rangle = \int_a^b h(t) \mathcal{X}_i(t) d(t)$$

Entonces, se define la medida de profundidad proyección aleatoria  $RPD$  para el  $i$ -ésimo dato como:

$$RPD_i = RPD(\mathcal{X}_i) = D(P_i^h)$$

Cuando se cuenta con una colección de proyecciones aleatorias  $\{h_l\}_{l=1}^M$ , la medida de profundidad se calcula usando todas las proyecciones. Con esto, si usamos un número suficiente de proyecciones aleatorias, el valor medio de las profundidades de los puntos proyectados define una profundidad para los datos funcionales. Para una exposición más detallada de la profundidad de proyección aleatoria véase [46].

La profundidad de proyección aleatoria de una curva  $X_i$  se define por:

- $RPD(\mathcal{X}_i) = \frac{1}{M} \sum_{l=1}^M D(P_i^{h_l})$  según Cuevas et al.[46].
- $RTD(\mathcal{X}_i) = \min_l D(P_I^{h_l})$  como se muestra en Cuesta-Albertos y Nieto-Reyes [47].

## CAPÍTULO 4

Como se indica en el capítulo 2, la fase de monitorización se ocupa de controlar el proceso, con el fin de detectar rápidamente cambios a partir de una muestra de calibrado o estabilizada. De acuerdo a la metodología establecida en el control de la calidad, un gráfico de control puede ser comprendido como un test de hipótesis que compara la media de la muestra calibrada con la media de muestra de control, lo cual es de gran utilidad porque en el diseño del gráfico de control nos permitirá calibrar la tasa de falsas alarmas (error tipo I) y tamaño de las muestras, con la finalidad de medir el rendimiento de la fase de monitorización (potencia).

En este contexto, la fase de monitorización del gráfico basado en la distancia de Mahalanobis hará posible el monitoreo de la igualdad de medias funcionales entre la muestra de monitoreo y una muestra estable. Como se menciona en [13], la carta de control basada en la distancia de Mahalanobis sigue siendo uno de los métodos más aplicados en el control estadístico porque se monitorea la similitud entre dos variables aleatorias tomando en cuenta la correlación para indicar la calidad de un único proceso productivo.

Para la carta de control que se propone en este trabajo, se considerará el estadístico basado en la distancia de Mahalanobis definido en la sección 3.4, y las distancias de la familia de Minkowski  $L^1$  y  $L^2$  presentadas en la sección 3.4.4. Esta metodología es de gran aporte ya que hace posible la monitorización de un proceso de variables aleatorias que toman valores en espacios de Hilbert. Es decir, extiende los conceptos estadísticos y del control de la calidad del campo multivariante al funcional; abordando la carta de control no solo desde el punto de vista de la producción, sino que también en el marco general de la detección de anomalías para datos funcionales.

En la metodología para el diseño de la fase de monitorización del gráfico de control basado en la distancia de Mahalanobis  $L^1$  estandarizada y  $L^2$  estandarizada, se determinará el límite de control superior (percentil) a partir de cual se determinará si la media de la muestra bajo control es estadísticamente igual a la media de la muestra de monitoreo. Se usará la técnica de remuestreo bootstrap suavizado para datos funcionales (véase sección 3.5.1) que a diferencia del bootstrap clásico, en su procedimiento toma en cuenta la estructura de covarianzas de los datos. También se aplicará la técnica de permutaciones descrita en la sección 3.5.2 y aplicada desde el punto de vista inferencial en [27].

En este capítulo se presentará la propuesta metodológica para la implementación de la fase de monitorización del gráfico de control basado en la distancia de Mahalanobis funcional,  $L^1$  estandarizada y  $L^2$  estandarizada que hace uso de las metodologías estadísticas funcionales (FDA) descritas en el capítulo 3. Para esto, se utilizará el software estadístico de código abierto R. Es importante mencionar que la facilidad de uso de herramientas FDA se debe al desarrollo de diversos paquetes implementados en R tales como *fda*, *fda.usc*, *fdahotelling*, *gmdf*, los mismos que serán utilizados en la presente tesis.

## 4.1 Metodología

En esta sección se presenta la metodología propuesta para monitorear y controlar procesos donde los datos son funciones. Se considerará una muestra bajo control (sin datos atípicos causados por fuentes variable aleatorias) y una muestra de monitoreo. El objetivo de dicha metodología es determinar si las nuevas muestras tomadas en tiempo real están o no bajo control.

En esta sección se desarrolla la metodología para la fase de monitorización de un gráfico de control de un variable aleatoria funcional  $\mathcal{X}$ , que toma valores en un espacio funcional  $\mathcal{E} = L^2$ , con  $T \subset \mathbb{R}$ . A partir de observaciones de la variable aleatoria funcional  $\mathcal{X}$  se cuenta con una muestra que está bajo control y otra de monitorización que son conjuntos de datos funcionales de tamaño  $n$  y  $m$  respectivamente, lo cual,

permitirá construir la fase de monitorización.

## 4.2 Procedimiento de construcción de un gráfico de control para la fase de monitorización

La fase de monitorización busca detectar rápidamente cambios o desviación en la distribución a partir de una muestra estable, lo que se logra recolectando datos preferentemente en tiempo real y probando hipótesis de igualdad de medias en el proceso. Utilizando la muestra bajo control, se puede estimar el percentil (límite de control superior) a un nivel  $\alpha$  de significancia de la distribución del estadístico basado en la distancia de Mahalanobis para dos muestras,  $L^1$  y  $L^2$  estandarizadas.

Para determinar el percentil se hará uso de las técnicas de remuestreo bootstrap suavizado y permutaciones. Inicialmente se propone el uso de medidas de profundidades para limpiar la muestra de posibles curvas atípicas, y remuestrear bajo la hipótesis nula, de acuerdo a lo expuesto por Febrero et al. en su trabajo de detección de anomalías para datos funcionales usando medidas de profundidad en [32].

La prueba de hipótesis que se aplica contrasta si existen desviaciones de la media entre la muestra de monitorizado  $\{\mathcal{X}_1(t), \mathcal{X}_2(t), \dots, \mathcal{X}_m(t)\}$  y la muestra de calibrado o bajo control  $\{\mathcal{X}_1(t), \dots, \mathcal{X}_n(t)\}$ .

En esta etapa los métodos de construcción de gráficos de control se basan en el contraste de hipótesis:

$$\begin{aligned} H_0 : \mu_F &= \mu_G \\ H_1 : \mu_F &\neq \mu_G \end{aligned} \tag{4.1}$$

Como se ha mencionado, en la fase de monitorización se utilizan las curvas correspondientes a la muestra de calibrado,  $\{\mathcal{X}_1(t), \dots, \mathcal{X}_n(t)\}$ , con la finalidad de detectar desviaciones con respecto al comportamiento del proceso bajo control. Para esto se seleccionan curvas del proceso, preferentemente en tiempo real  $\{\mathcal{X}_1(t), \dots, \mathcal{X}_m(t)\}$ , y se contrasta la hipótesis 4.1.

En esta fase, se emplea la longitud media de racha (ARL) para evaluar el desempeño del gráficos de control. Bajo este contexto, las pruebas de hipótesis que se aplica contrasta si existen desviaciones de la media entre los datos obtenidos en la fase de monitorización, también llamados muestra de monitorizado y los datos de referencia o muestra de calibrado.

A continuación, se detalla la metodología del diseño de gráfico de control para la fase de monitorización, considerando la técnica de remuestreo bootstrap suavizado y permutaciones.

1. Agrupar la muestra de calibrado  $\{\mathcal{X}_1(t), \dots, \mathcal{X}_n(t)\}$  y monitoreo  $\{\mathcal{X}_1(t), \mathcal{X}_2(t), \dots, \mathcal{X}_m(t)\}$ , a la muestra agrupada la llamaremos  $\mathcal{Z}$ .
2. Calcular las profundidades  $D(\mathcal{Z}_i)_{i=1}^{m+n}$ ; y, para determinar el límite de control superior (LCS) se sigue el procedimiento a continuación:
  - Obtener  $B$  remuestras bootstrap de tamaño  $m + n$  de las curvas obtenidas después del recorte del  $\alpha\%$  de las curvas menos profundas. Sean las muestras bootstrap  $\mathcal{Z}_i^{*b}$  con  $i = 1, \dots, n + m$ ,  $b = 1, \dots, B$  se obtendrán como sigue:
    - Se realiza un muestreo uniforme, con  $i^*$  de  $1, \dots, [(n + m)(1 - \alpha)]$ , con reemplazo.
    - Se genera  $V_{i^*}$  como un proceso gaussiano con media cero y matriz de varianza y covarianza  $\gamma \Sigma_{\mathcal{Z}}$ , con  $\gamma \in [0, 1]$ . Donde  $\Sigma_{\mathcal{Z}}$  es la matriz de varianzas y covarianzas de las observaciones  $\mathcal{Z}_1, \dots, \mathcal{Z}_{[(n+m)(1-\alpha)]}$
    - Finalmente se obtiene  $\mathcal{Z}_i^{*b} = \mathcal{Z}_{i^*} + V_{i^*}$
  - Reagrupar  $\mathcal{Z}_i^{*b}$  en las muestras  $\mathcal{X}'$  y  $\mathcal{Y}'$  de tamaño  $m$  y  $n$  respectivamente.

- Calcular el estadístico  $W$  y el  $LCS$  como el percentil de la distribución de  $W(\mathcal{X}', \mathcal{Y}')$ , donde  $W$  es el estadístico 3.14, 3.16 o 3.15
3. Se procede a monitorizar el proceso. Si se observa que  $W(\mathcal{X}', \mathcal{Y}') \geq LCS$ , entonces el proceso está fuera de control
1. Agrupar la muestra de calibrado  $\{\mathcal{X}_1(t), \dots, \mathcal{X}_n(t)\}$  y monitoreo  $\{\mathcal{X}_1(t), \mathcal{X}_2(t), \dots, \mathcal{X}_m(t)\}$ , a la muestra agrupada la llamaremos  $\mathcal{Z}$ .
  2. Calcular las profundidades  $D(\mathcal{Z}_i)_{i=1}^{m+n}$ ; y, para determinar el límite de control superior (LCS) se sigue el procedimiento a continuación:
    - Considerar todas las posibles permutaciones de la muestra conjunta recortada  $\mathcal{Z}$  y reasignar aleatoriamente las curvas a los grupos  $\mathcal{X}'$  y  $\mathcal{Y}'$
    - Calcular el estadístico  $W$  y el  $LCS$  como el percentil de la distribución de  $W(\mathcal{X}', \mathcal{Y}')$ , donde  $W$  es el estadístico 3.14, 3.16 o 3.15
  3. Se procede a monitorizar el proceso. Si se observa que  $W(\mathcal{X}', \mathcal{Y}') \geq LCS$ , entonces el proceso está fuera de control

### 4.3 Estudio de Simulación

El rendimiento de un gráfico de control generalmente se mide por su potencia  $(1 - \beta)$ , definida como la probabilidad de detectar una señal de fuera de control cuando el proceso efectivamente no está bajo control. En cambio, cuando el proceso está bajo control, la probabilidad de detectar una observación fuera de control se denomina error tipo I o tasa de falsas alarmas.

Otra medida de desempeño del gráfico de control es el ARL (Average Run Length), que se define como el número promedio de puntos que se deben graficar en una carta de control antes de que un punto indique una salida de control. Si las observaciones del proceso no se encuentran correlacionadas, el ARL puede ser calculado así:  $ARL = \frac{1}{p}$ , donde  $p$  es la probabilidad de que cualquier punto esté fuera de los límites de control.



Se identifican dos tipos de ARL

ARL bajo control ( $ARL_0$ ): Que es el número esperado de puntos o muestras hasta que el gráfico de control de una señal de fuera de control, bajo la condición de que el proceso actual se encuentra bajo control. Se calcula como  $\frac{1}{\alpha}$ , siendo  $\alpha$  la probabilidad de cometer el error tipo I.

ARL fuera de control ( $ARL_1$ ): Que es el número esperado de puntos o muestras hasta que el gráfico de control dé una señal de fuera de control, bajo la condición de que el proceso actual se encuentra fuera de control. Se calcula como  $\frac{1}{1-\beta}$ , siendo  $\beta$  la probabilidad de cometer el error tipo II.

Es claro que se desea que las falsas alarmas de un proceso que está fuera de control sean menos frecuentes. A menor  $\alpha$  mayor  $ARL_0$ . Así mismo, necesitamos detectar pronto cuando el proceso está fuera de control por eso necesitamos un  $\beta$  pequeño.

Se diseñan diferentes escenarios de media y covarianza. Las integrales involucradas en el cálculo de los estadísticos a monitorizarse en el gráfico de control se obtienen usando la regla de la cuadratura del rectángulo basada en una cuadrícula uniforme de 100 puntos.

Como no se conoce la distribución  $F$  para calcular la potencia de los gráficos de control se diseña una simulación de Monte Carlo. En [32] se consideran datos funcionales  $\mathcal{X}_1, \dots, \mathcal{X}_n$  obtenidos como realizaciones de un proceso estocástico, asumiendo trayectorias continuas en el período de observación  $T = [a, b] = [0, 1]$  e independencia entre curvas. Se generan realizaciones de un proceso estocástico gaussiano como sigue

$$\mathcal{X}(t) = \mu(t) + \epsilon(t) \quad (4.2)$$

donde

$$\mu(t) = \mathbb{E}(\mathcal{X}(t)) = 30t(1-t)^{\frac{3}{2}} \quad (4.3)$$

también  $\epsilon(t)$  es un proceso con media cero y operadores de covarianza que se describen a continuación

**Escenarios de Operadores de Covarianza:** el operador de covarianza se diseña para simular trayectorias diferenciables en escenarios estacionarios y no estacionarios

**Escenario A:** para las trayectorias estacionarias se usará el siguiente operador

$$(\mathcal{K}g)(t) = \int_0^1 e^{-2(s-t)^2} g(s) ds \quad \forall t \in [0, 1] \quad (4.4)$$

**Escenario B:** para las trayectorias no estacionarias se usará el siguiente operador

$$(\mathcal{K}g)(t) = \int_0^1 e^{-2(s-t)^2} (s+0.5)(t+0.5)g(s) ds \quad \forall t \in [0, 1] \quad (4.5)$$

También se plantean los siguientes escenarios para la media

**Cambio de la media en desplazamiento y forma:** el gráfico de control está diseñado para monitorizar la media funcional, para detectar cambios en magnitud y forma se diseñan los siguientes escenarios

**Escenario 1:** para el cambio en desplazamiento se plantea la media 4.6 que varía de acuerdo al parámetro  $\delta \in \{0, 0.1, \dots, 1\}$

$$\mu(t) = 30t \cdot (1-t)^{\frac{3}{2}} + \delta \quad (4.6)$$

**Escenario 2:** para la variación de la media en forma se tiene 4.7 cuyo parámetro de ajuste es  $\eta \in \{0, 0.1, \dots, 0.6\}$

$$\mu(t) = (1 - \eta)30t \cdot (1 - t)^{\frac{3}{2}} + \eta \cdot 30t^{\frac{3}{2}}(1 - t) \quad (4.7)$$

A continuación, en la figura 4.1 en el panel a) y b) se muestran las medias funcionales, con cambios en desplazamiento y forma respectivamente.

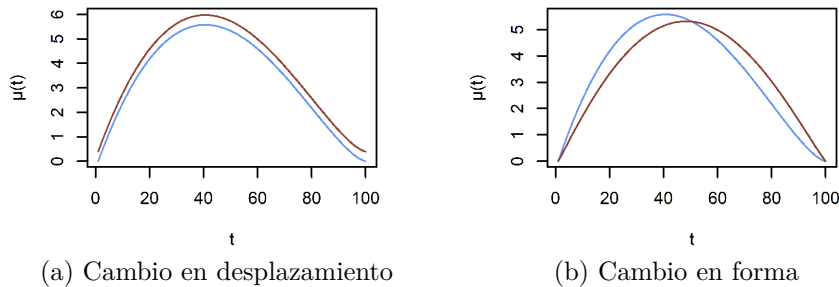


Figura 4.1: En (a) vemos un ejemplo del desplazamiento de la media con  $\delta = 0.2$ , mientras en (b) vemos un ejemplo de cambio en la media de forma con  $\eta = 0.4$

Los diferentes escenarios 1A, 2A, 1B, 2B permitirán evaluar el rendimiento de la fase de monitorización del gráfico de control para cambios en la media en forma y en desplazamiento considerando los operadores 4.4 y 4.5.

### 4.3.1 Medición del rendimiento del gráfico de monitorización

Se compara el desempeño del gráfico de control para la fase de monitorización a partir de la generación de las muestras de calibrado y monitoreo de tamaño 25, 50, 100. Al unir las dos muestras se recorta el 2.5% de las curvas atípicas basándonos en las medidas de profundidad definidas en la sección 3.6. A partir de los procedimientos bootstrap suavizado (con factor de suavizamiento  $\gamma = 0.05$ ) y permutaciones se establece el límite de control superior que es el percentil de los estadísticos a un nivel de significancia  $\alpha = 0.05$ . Siguiendo este procedimiento se estiman la tasa de

falsas alarmas (error tipo I) cuando el proceso está bajo control y la potencia de la prueba cuando está fuera de control. Para calcular el error tipo I (cuando  $\delta = 0$  ó  $\eta = 0$ ) y desempeño del gráfico de monitorización simulamos  $M = 2000$  pares de muestras aleatorias independientes de tamaño 25, 50, 100 considerando los escenarios descritos en la sección 4.3.

En las tablas 4.1, 4.2, 4.3, 4.4, 4.5, 4.6 se presentan los resultados de los valores empíricos de las tasas de falsas alarmas y desempeño del gráfico de control (potencia) para detectar un cambio de la media en desplazamiento considerando el operador de trayectorias estacionarias.

En la tabla 4.1 se presentan los resultados de las tasas de falsas alarmas y la potencia correspondiente al escenario que ha sido diseñado para detectar cambios de la media en desplazamiento, y, el operador de covarianza estacionario. En esta tabla se observa que la tasas de falsas alarmas subestiman el valor nominal  $\alpha = 0.05$ . Si bien es cierto, se alcanza la potencia igual a 1 cuando  $\delta = 1.1$ , sin embargo, los gráficos de control que monitorizan los estadísticos  $L^1$  y  $L^2$  convergen más rápido que el gráfico de control basado en la distancia de Mahalanobis. También se ve que tanto los valores estimados por el método de bootstrap y permutaciones son cercanos a los valores reales (estimados por Montecarlo).

En la tabla 4.2 se presentan los resultados de las tasas de falsas alarmas y potencia cuando  $n = 25$  y  $m = 50$ . En este escenario la tasa de falsas alarmas subestima el valor nominal  $\alpha = 0.05$  y la potencia igual a uno se alcanza cuando  $\delta = 1.1$ ; además, los valores estimados por el método de bootstrap como permutaciones son cercanos a los valores estimados por Montecarlo.

En la tabla 4.3 se presentan los resultados de las tasas de falsas alarmas y potencia cuando  $n = 25$  y  $m = 100$ . En este escenario se la tasa de falsas alarmas converge al valor nominal  $\alpha = 0.05$ . Los valores estimados de la potencia converge a 1 cuando  $\delta = 1$ . Se tiene desempeño similar para los tres estadísticos; además, los valores estimados tanto por el método bootstrap como permutaciones son cercanos a los estimados por

Montecarlo.

En la tabla 4.4 se presentan los resultados de las tasas de falsas alarmas y potencia cuando  $n = 50$  y  $m = 50$ . Para este escenario se observa que se tiene un mejor desempeño en detectar cambios de la media en desplazamiento cuando se consideran los estadísticos  $L^1$  y  $L^2$ . También, se ve que la tasa de falsas alarmas converge al valor nominal  $\alpha = 0.05$ .

En la tabla 4.5 se presentan los resultados de las tasas de falsas alarmas y potencia cuando  $n = 50$  y  $m = 100$ . En este escenario se observa que se tiene desempeño similar para los tres estadísticos para detectar cambios de desplazamiento de la media. También, se ve que la tasa de falsas alarmas converge a  $\alpha$ .

En la tabla 4.6 se presentan los resultados de las tasas de falsas alarmas y potencia cuando  $n = 100$  y  $m = 100$ . En este escenario se ve que en general se tiene mejor desempeño cuando se consideran los estadísticos  $L^1$  y  $L^2$  estandarizados. La tasa de falsas alarmas converge a  $\alpha = 0.05$  para los tres estadísticos; además, tanto el método de bootstrap como permutaciones se obtienen valores cercanos a los estimados por el método de Montecarlo.

En la tabla 4.7 se presentan los resultados del escenario 1B (cambio de la media en desplazamiento y operador de covarianza no estacionario). Se observa que para el estadístico basado en la distancia de Mahalanobis la potencia converge a 1 cuando  $\delta = 0.5$ , mientras que para los estadísticos  $L^1$  estandarizado y  $L^2$  estandarizado cuando  $\delta = 0.8$ . También se observa que el método de bootstrap genera valores estimados más cercanos de la potencia calculada por Montecarlo. Respecto a la tasa de falsas alarmas (cuando  $\delta = 0$ ) se subestima, considerando que el valor nominal es  $\alpha = 0.05$ .

En la tabla 4.8 se presentan los resultados de tasas de falsas alarmas (error tipo I) y potencia (error tipo II) para  $n = 25$  y  $m = 50$ . Para el estadístico basado en la distancia de Mahalanobis, se observa una convergencia a 1 cuando  $\delta = 0.4$ , y, el error tipo I converge al valor nominal  $\alpha = 0.05$ ; para los estadísticos  $L^1$  estandarizado y

$L^2$  estandarizado, la potencia converge a 1 cuando  $\delta = 0.8$ , además el error tipo I se subestima.

De igual forma que en la tabla 4.7 el estadístico basado en la distancia de Mahalanobis tiene mejor potencia que los otros dos estadísticos, y la potencia considerando el método bootstrap es más cercana a la potencia estimada por Montecarlo.

En la tabla 4.9 se muestran los resultados del error tipo I y potencia de la fase de monitorización del gráfico de control para  $n = 25$  y  $m = 100$ . Se observa que para el estadístico basado en la distancia de Mahalanobis, las tasa de falsas de alarmas converge al valor nominal ( $\alpha = 0.05$ ), mientras que para los otros dos estadísticos dicha tasa se subestima.

La potencia estimada con el estadístico basado en la distancia de Mahalanobis, converge a 1 cuando  $\delta = 0.4$ , mientras que para los estadísticos  $L^1$  estandarizado y  $L^2$  estandarizado, la potencia converge a 1 cuando  $\delta = 0.7$ . Además, cabe notar que tanto la tasa de falsas alarmas como la potencia estimados a partir del método bootstrap son más cercados a los valores calculados por Montecarlo.

En la tabla 4.10 se presentan los resultados de las tasas de falsas alarmas y potencia para  $n = 50$  y  $m = 50$ . En este escenario se considera el cambio en desplazamiento de la media y el operador no estacionario de covarianza. Se observa que las tasas de falsas alarmas convergen al valor nominal  $\alpha = 0.05$ . Además, la potencia del gráfico de control basado en la distancia de Mahalanobis converge más rápido a 1 (cuando  $\delta = 0.4$ ) que los otros dos gráficos de control que monitorizan los estadísticos  $L^1$  y  $L^2$ . Los valores estimados por el método de bootstrap y permutaciones son cercanos a los estimados por Montecarlo.

En la tabla 4.11 se presentan los resultados del error tipo I y potencia para  $n = 50$  y  $m = 100$ . Se observa que la tasa de falsas alarmas converge al valor nominal  $\alpha = 0.05$ . La potencia del gráfico de control basado en la distancia de Mahalanobis es más rápida (cuando  $\delta = 0.3$ ) que los gráficos de control que monitorizan los estadísticos

$L^1$  y  $L^2$ . Tanto los valores estimados por bootstrap y permutaciones son cercanos a los estimados por Montecarlo.

En la tabla 4.12 se presentan los resultados de tasas de falsas alarmas y potencia para  $n = 100$  y  $m = 100$ . Se observa que la tasa de falsas alarmas converge al valor nominal  $\alpha = 0.05$ . La potencia del gráfico de control basado en la distancia de Mahalanobis es más rápida (cuando  $\delta = 0.3$ ) que los gráficos de control que monitorizan los estadísticos  $L^1$  y  $L^2$ . Tanto los valores estimados por bootstrap y permutaciones son cercanos a los estimados por Montecarlo.

A continuación se analizará el escenarios 2A, considerando el estadístico basado en la distancia de Mahalanobis  $L^1$  estandarizado y  $L^2$  estandarizado y también variación de tamaños de la muestra de monitoreo y calibrado.

En la tabla 4.13 se presentan los resultados de las tasas de falsas alarmas y potencia para gráfico de control que monitoriza el estadístico basado en la distancia de Mahalanobis. En este escenario se ve que la tasa de falsas alarmas convergen a  $\alpha = 0.05$ . Se observa que mejora el desempeño de detectar cambios en forma de la media a medida que aumenta el tamaño de las muestras.

En la tabla 4.14 se presentan los resultados de las tasas de falsas alarmas y potencia para gráfico de control que monitoriza el estadístico  $L^1$  estandarizado. En este escenario se ve que la tasa de falsas alarmas convergen a  $\alpha = 0.05$ . Se observa que mejora el desempeño de detectar cambios en forma de la media a medida que aumenta el tamaño de las muestras.

En la tabla 4.15 se presentan los resultados de las tasas de falsas alarmas y potencia para gráfico de control que monitoriza el estadístico  $L^2$  estandarizado. En este escenario se ve que la tasa de falsas alarmas convergen a  $\alpha = 0.05$ . Se observa que mejora el desempeño de detectar cambios en forma de la media a medida que aumenta el tamaño de las muestras.

En la tabla 4.16 se presentan los resultados de las tasas de falsas alarmas y potencia para gráfico de control que monitoriza el estadístico basado en la distancia de Mahalanobis. En este escenario se ve que la tasa de falsas alarmas convergen a  $\alpha = 0.05$ . Tiene una convergencia a la potencia más rápido que los gráficos  $L^1$  y  $L^2$ .

En la tabla 4.17 se presentan los resultados de las tasas de falsas alarmas y potencia para gráfico de control que monitoriza el estadístico basado  $L^1$  estandarizado. En este escenario se ve que la tasa de falsas alarmas convergen a  $\alpha = 0.05$ . Se observa que mejora el desempeño de detectar cambios en forma de la media a medida que aumenta el tamaño de las muestras.



Tabla 4.1: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 1A para  $n=25$  y  $m=25$

Método	Muestra 1	Muestra 2	Estadístico	Potencia											
				0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1
Montecarlo				0,034	0,050	0,066	0,132	0,246	0,358	0,506	0,676	0,782	0,882	0,958	1,000
Bootstrap			Mahalanobis	0,042	0,060	0,078	0,148	0,262	0,396	0,528	0,694	0,808	0,902	0,964	1,000
Permutaciones				0,040	0,054	0,062	0,126	0,244	0,344	0,510	0,686	0,794	0,898	0,964	1,000
Montecarlo				0,046	0,078	0,152	0,282	0,420	0,592	0,772	0,856	0,906	0,974	0,998	1,000
Bootstrap	25	25	$L^1$ estandarizado	0,044	0,076	0,140	0,280	0,412	0,576	0,752	0,848	0,894	0,964	0,994	1,000
Permutaciones				0,040	0,050	0,104	0,218	0,376	0,524	0,714	0,834	0,880	0,954	0,990	1,000
Montecarlo				0,052	0,074	0,164	0,288	0,424	0,604	0,774	0,866	0,900	0,972	0,998	1,000
Bootstrap			$L^2$ estandarizado	0,046	0,058	0,126	0,256	0,394	0,552	0,720	0,846	0,882	0,954	0,992	1,000
Permutaciones				0,038	0,040	0,096	0,192	0,344	0,486	0,680	0,816	0,852	0,944	0,982	1,000

Escenario 1A

Tabla 4.2: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 1A para  $n = 25$  y  $m = 50$

Método	Muestra 1	Muestra 2	Estadístico	Potencia											
				0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1
Montecarlo				0,046	0,056	0,070	0,156	0,270	0,430	0,626	0,802	0,876	0,960	0,990	1,000
Bootstrap			Mahalanobis	0,044	0,056	0,076	0,158	0,280	0,430	0,634	0,802	0,874	0,960	0,988	1,000
Permutaciones				0,046	0,066	0,102	0,200	0,302	0,510	0,686	0,846	0,904	0,966	0,994	1,000
Montecarlo				0,026	0,038	0,078	0,172	0,332	0,520	0,718	0,862	0,936	0,980	0,996	1,000
Bootstrap	25	50	$L^1$ estandarizado	0,032	0,052	0,096	0,196	0,368	0,556	0,740	0,876	0,942	0,980	0,996	1,000
Permutaciones				0,046	0,092	0,162	0,314	0,488	0,680	0,828	0,936	0,984	0,996	1,000	1,000
Montecarlo				0,020	0,038	0,070	0,174	0,316	0,500	0,684	0,844	0,922	0,974	0,996	1,000
Bootstrap			$L^2$ estandarizado	0,022	0,046	0,078	0,176	0,336	0,516	0,700	0,846	0,934	0,974	0,996	1,000
Permutaciones				0,058	0,090	0,144	0,278	0,470	0,662	0,808	0,928	0,980	0,992	0,998	1,000

Escenario 1A

Tabla 4.3: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 1A para  $n = 25$  y  $m = 100$

Método	Muestra 1	Muestra 2	Estadístico	Potencia											
				0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1	1.1
Montecarlo				0,042	0,058	0,104	0,216	0,390	0,584	0,730	0,906	0,980	0,994	0,998	1,000
Bootstrap			Mahalanobis	0,042	0,058	0,098	0,208	0,388	0,580	0,724	0,896	0,978	0,992	0,998	1,000
Permutaciones				0,046	0,074	0,144	0,260	0,440	0,648	0,784	0,930	0,994	0,996	0,998	1,000
Montecarlo				0,032	0,030	0,076	0,202	0,384	0,598	0,740	0,908	0,964	0,986	0,996	1,000
Bootstrap	25	100	$L^1$ estandarizado	0,026	0,032	0,080	0,218	0,398	0,614	0,762	0,910	0,964	0,986	0,996	1,000
Permutaciones				0,056	0,176	0,338	0,524	0,736	0,866	0,940	0,984	1,000	1,000	1,000	1,000
Montecarlo				0,042	0,022	0,070	0,192	0,370	0,564	0,718	0,892	0,962	0,984	0,996	1,000
Bootstrap			$L^2$ estandarizado	0,048	0,022	0,066	0,182	0,354	0,550	0,714	0,884	0,956	0,982	0,996	1,000
Permutaciones				0,046	0,172	0,314	0,514	0,724	0,860	0,932	0,986	1,000	1,000	1,000	1,000

Escenario 1A

Tabla 4.4: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 1A para  $n = 50$  y  $m = 50$

Método	Muestra 1	Muestra 2	Estadístico	Potencia									
				0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
Montecarlo				0,050	0,066	0,136	0,294	0,524	0,714	0,910	0,970	0,992	1,000
Bootstrap			Mahalanobis	0,054	0,064	0,132	0,294	0,524	0,716	0,900	0,968	0,992	1,000
Permutaciones				0,060	0,062	0,124	0,280	0,532	0,714	0,904	0,970	0,994	1,000
Montecarlo				0,056	0,086	0,238	0,420	0,670	0,838	0,938	0,986	1,000	1,000
Bootstrap	50	50	$L^1$ estandarizado	0,054	0,092	0,238	0,434	0,672	0,834	0,938	0,982	1,000	1,000
Permutaciones				0,052	0,074	0,200	0,374	0,618	0,788	0,918	0,980	1,000	1,000
Montecarlo				0,054	0,092	0,226	0,420	0,658	0,830	0,932	0,982	1,000	1,000
Bootstrap			$L^2$ estandarizado	0,054	0,086	0,218	0,410	0,652	0,806	0,918	0,980	1,000	1,000
Permutaciones				0,052	0,066	0,186	0,344	0,578	0,774	0,894	0,970	1,000	1,000

Escenario 1A

Tabla 4.5: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 1A para  $n = 50$  y  $m = 100$

Método	Muestra 1	Muestra 2	Estadístico	Potencia								
				0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8
Montecarlo				0,052	0,056	0,180	0,404	0,644	0,878	0,962	1,000	1,000
Bootstrap			Mahalanobis	0,042	0,046	0,162	0,384	0,610	0,866	0,956	1,000	1,000
Permutaciones				0,056	0,064	0,188	0,428	0,660	0,896	0,966	1,000	1,000
Montecarlo				0,038	0,050	0,200	0,408	0,684	0,882	0,954	1,000	1,000
Bootstrap	50	100	$L^1$ estandarizado	0,042	0,062	0,230	0,440	0,710	0,900	0,958	1,000	1,000
Permutaciones				0,040	0,118	0,300	0,604	0,792	0,940	0,978	1,000	1,000
Montecarlo				0,046	0,048	0,172	0,370	0,662	0,862	0,942	0,994	1,000
Bootstrap			$L^2$ estandarizado	0,046	0,056	0,194	0,400	0,674	0,878	0,952	0,996	1,000
Permutaciones				0,044	0,112	0,284	0,558	0,772	0,938	0,974	0,996	1,000

Escenario 1A

Tabla 4.6: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1-\beta$  en el escenario 1A para  $n = 100$  y  $m = 100$

Método	Muestra 1	Muestra 2	Estadístico	Potencia							
				0	0.1	0.2	0.3	0.4	0.5	0.6	0.7
Montecarlo				0	0,106	0,260	0,620	0,864	0,980	0,996	1,000
Bootstrap			Mahalanobis	0	0,106	0,250	0,614	0,862	0,978	0,996	1,000
Permutaciones				0	0,110	0,254	0,598	0,862	0,978	0,996	1,000
Montecarlo				0	0,114	0,348	0,702	0,910	0,980	0,998	1,000
Bootstrap	100	100	$L^1$ estandarizado	0	0,136	0,398	0,746	0,928	0,988	0,998	1,000
Permutaciones				0	0,130	0,366	0,712	0,916	0,982	0,998	1,000
Montecarlo				0	0,116	0,362	0,706	0,908	0,982	0,998	1,000
Bootstrap			$L^2$ estandarizado	0	0,126	0,370	0,738	0,918	0,986	0,998	1,000
Permutaciones				0	0,116	0,352	0,692	0,902	0,978	0,998	1,000

Escenario 1A

Tabla 4.7: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 1B para  $n = 25$  y  $m = 25$

Método	Muestra 1	Muestra 2	Estadístico	Potencia										
				0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Montecarlo				0,030	0,126	0,466	0,852	0,990	1,000	1,000	1,000	1,000	1,000	1,000
Bootstrap			Mahalanobis	0,042	0,136	0,482	0,864	0,990	1,000	1,000	1,000	1,000	1,000	1,000
Permutaciones				0,032	0,130	0,472	0,874	0,990	1,000	1,000	1,000	1,000	1,000	1,000
Montecarlo				0,038	0,092	0,180	0,320	0,546	0,690	0,866	0,988	1,000	1,000	1,000
Bootstrap	25	25	$L^1$ estandarizado	0,036	0,094	0,174	0,318	0,538	0,682	0,866	0,988	1,000	1,000	1,000
Permutaciones				0,022	0,064	0,148	0,264	0,480	0,636	0,812	0,976	1,000	1,000	1,000
Montecarlo				0,050	0,082	0,190	0,338	0,582	0,758	0,924	0,996	1,000	1,000	1,000
Bootstrap			$L^2$ estandarizado	0,038	0,078	0,180	0,324	0,548	0,720	0,914	0,994	1,000	1,000	1,000
Permutaciones				0,026	0,060	0,136	0,282	0,514	0,708	0,888	0,992	1,000	1,000	1,000

Escenario 1B

Tabla 4.8: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 1B para  $n = 25$  y  $m = 50$

Método	Muestra 1	Muestra 2	Estadístico	Potencia										
				0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Montecarlo				0,046	0,140	0,560	0,928	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Bootstrap			Mahalanobis	0,048	0,140	0,576	0,928	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Permutaciones				0,056	0,166	0,598	0,942	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Montecarlo				0,022	0,036	0,086	0,238	0,484	0,666	0,856	0,982	1,000	1,000	1,000
Bootstrap	25	50	$L^1$ estandarizado	0,024	0,048	0,116	0,276	0,518	0,682	0,874	0,982	1,000	1,000	1,000
Permutaciones				0,050	0,104	0,226	0,414	0,678	0,838	0,946	0,996	1,000	1,000	1,000
Montecarlo				0,020	0,036	0,096	0,250	0,528	0,724	0,922	0,994	0,998	1,000	1,000
Bootstrap			$L^2$ estandarizado	0,022	0,042	0,102	0,268	0,528	0,724	0,922	0,994	0,998	1,000	1,000
Permutaciones				0,046	0,102	0,228	0,436	0,708	0,866	0,968	0,998	1,000	1,000	1,000

Escenario 1B



Tabla 4.9: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 1B para  $n = 25$  y  $m = 100$

Método	Muestra 1	Muestra 2	Estadístico	Potencia										
				0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Montecarlo				0,050	0,170	0,656	0,984	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Bootstrap			Mahalanobis	0,052	0,174	0,648	0,984	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Permutaciones				0,064	0,202	0,700	0,986	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Montecarlo				0,024	0,036	0,114	0,274	0,512	0,718	0,872	0,990	0,998	1,000	1,000
Bootstrap	25	100	$L^1$ estandarizado	0,030	0,044	0,114	0,276	0,516	0,726	0,876	0,990	0,998	1,000	1,000
Permutaciones				0,138	0,182	0,364	0,608	0,834	0,936	0,974	1,000	1,000	1,000	1,000
Montecarlo				0,022	0,036	0,108	0,272	0,546	0,778	0,934	1,000	1,000	1,000	1,000
Bootstrap			$L^2$ estandarizado	0,022	0,036	0,096	0,240	0,516	0,762	0,924	1,000	1,000	1,000	1,000
Permutaciones				0,132	0,174	0,378	0,656	0,892	0,964	0,992	1,000	1,000	1,000	1,000

Escenario 1B

Tabla 4.10: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 1B para  $n = 50$  y  $m = 50$

Método	Muestra 1	Muestra 2	Estadístico	Potencia										
				0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Montecarlo				0,054	0,288	0,850	0,998	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Bootstrap			Mahalanobis	0,060	0,282	0,848	0,998	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Permutaciones				0,052	0,252	0,836	0,996	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Montecarlo				0,052	0,110	0,248	0,520	0,810	0,940	0,980	1,000	1,000	1,000	1,000
Bootstrap	50	50	$L^1$ estandarizado	0,052	0,110	0,250	0,538	0,814	0,940	0,980	1,000	1,000	1,000	1,000
Permutaciones				0,052	0,110	0,238	0,520	0,774	0,934	0,978	1,000	1,000	1,000	1,000
Montecarlo				0,052	0,116	0,288	0,594	0,874	0,972	0,994	1,000	1,000	1,000	1,000
Bootstrap			$L^2$ estandarizado	0,052	0,100	0,268	0,582	0,866	0,964	0,994	1,000	1,000	1,000	1,000
Permutaciones				0,052	0,092	0,238	0,546	0,812	0,960	0,994	1,000	1,000	1,000	1,000

Escenario 1B

Tabla 4.11: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 1B para  $n = 50$  y  $m = 100$

Método	Muestra 1	Muestra 2	Estadístico	Potencia										
				0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Montecarlo				0,052	0,330	0,960	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Bootstrap			Mahalanobis	0,052	0,314	0,958	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Permutaciones				0,046	0,334	0,964	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Montecarlo				0,048	0,064	0,232	0,550	0,814	0,974	0,996	1,000	1,000	1,000	1,000
Bootstrap	50	100	$L^1$ estandarizado	0,042	0,094	0,258	0,584	0,852	0,978	0,998	1,000	1,000	1,000	1,000
Permutaciones				0,048	0,170	0,372	0,722	0,916	0,990	1,000	1,000	1,000	1,000	1,000
Montecarlo				0,042	0,330	0,960	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Bootstrap			$L^2$ estandarizado	0,042	0,314	0,958	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Permutaciones				0,046	0,334	0,964	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000

Escenario 1B

Tabla 4.12: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 1B para  $n = 100$  y  $m = 100$

Método	Muestra 1	Muestra 2	Estadístico	Potencia										
				0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1
Montecarlo				0,052	0,518	0,994	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Bootstrap			Mahalanobis	0,052	0,496	0,994	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Permutaciones				0,052	0,498	0,996	1,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000
Montecarlo				0,042	0,172	0,496	0,820	0,974	0,998	1,000	1,000	1,000	1,000	1,000
Bootstrap	100	100	$L^1$ estandarizado	0,050	0,178	0,518	0,838	0,982	0,998	1,000	1,000	1,000	1,000	1,000
Permutaciones				0,038	0,150	0,492	0,800	0,970	0,998	1,000	1,000	1,000	1,000	1,000
Montecarlo				0,044	0,174	0,540	0,868	0,988	1,000	1,000	1,000	1,000	1,000	1,000
Bootstrap			$L^2$ estandarizado	0,044	0,170	0,546	0,872	0,988	1,000	1,000	1,000	1,000	1,000	1,000
Permutaciones				0,048	0,130	0,508	0,852	0,988	1,000	1,000	1,000	1,000	1,000	1,000

Escenario 1B

Tabla 4.13: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 2A. Gráfico basado en la distancia de Mahalanobis

Muestra 1	Muestra 2	Método	Potencia									
			0	0.01	0.015	0.02	0.025	0.03	0.035	0.04	0.045	0.05
20	20	Montecarlo	0,040	0,084	0,196	0,390	0,602	0,784	0,898	0,970	0,998	1,000
		Bootstrap	0,040	0,090	0,204	0,406	0,610	0,800	0,906	0,972	0,998	1,000
		Permutaciones	0,034	0,080	0,194	0,356	0,578	0,768	0,884	0,960	0,994	1,000
20	40	Montecarlo	0,042	0,102	0,258	0,508	0,794	0,920	0,988	1,000	1,000	1,000
		Bootstrap	0,038	0,096	0,240	0,466	0,768	0,906	0,978	1,000	1,000	1,000
		Permutaciones	0,060	0,156	0,320	0,580	0,846	0,934	0,990	1,000	1,000	1,000
40	40	Montecarlo	0,048	0,252	0,508	0,814	0,962	1,000	1,000	1,000	1,000	1,000
		Bootstrap	0,046	0,244	0,504	0,796	0,954	1,000	1,000	1,000	1,000	1,000
		Permutaciones	0,044	0,244	0,500	0,798	0,958	1,000	1,000	1,000	1,000	1,000
40	60	Montecarlo	0,048	0,274	0,562	0,866	0,980	1,000	1,000	1,000	1,000	1,000
		Bootstrap	0,040	0,278	0,562	0,864	0,980	1,000	1,000	1,000	1,000	1,000
		Permutaciones	0,046	0,292	0,572	0,868	0,984	1,000	1,000	1,000	1,000	1,000
60	60	Montecarlo	0,052	0,338	0,712	0,944	0,998	1,000	1,000	1,000	1,000	1,000
		Bootstrap	0,050	0,314	0,708	0,942	0,998	1,000	1,000	1,000	1,000	1,000
		Permutaciones	0,040	0,332	0,708	0,942	0,998	1,000	1,000	1,000	1,000	1,000

Gráfico basado en la distancia de Mahalanobis

Tabla 4.14: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 2A. Gráfico basado en el estadístico  $L^1$

Muestra 1	Muestra 2	Método	Potencia						
			0	0.1	0.2	0.3	0.4	0.5	0.6
20	20	Montecarlo	0,040	0,032	0,056	0,110	0,324	0,704	1,000
		Bootstrap	0,048	0,056	0,106	0,184	0,572	0,914	1,000
		Permutaciones	0,042	0,040	0,072	0,154	0,476	0,850	1,000
20	40	Montecarlo	0,042	0,036	0,046	0,144	0,416	0,862	1,000
		Bootstrap	0,044	0,040	0,048	0,170	0,526	0,934	1,000
		Permutaciones	0,062	0,078	0,124	0,416	0,816	1,000	1,000
40	40	Montecarlo	0,054	0,054	0,142	0,566	0,996	1,000	1,000
		Bootstrap	0,054	0,058	0,142	0,590	0,996	1,000	1,000
		Permutaciones	0,050	0,044	0,124	0,514	0,996	1,000	1,000
40	60	Montecarlo	0,042	0,046	0,122	0,544	1,000	1,000	1,000
		Bootstrap	0,042	0,044	0,122	0,582	1,000	1,000	1,000
		Permutaciones	0,036	0,072	0,204	0,860	1,000	1,000	1,000
60	60	Montecarlo	0,046	0,074	0,268	0,966	1,000	1,000	1,000
		Bootstrap	0,052	0,082	0,292	0,994	1,000	1,000	1,000
		Permutaciones	0,040	0,068	0,246	0,950	1,000	1,000	1,000

Gráfico basado en el estadístico  $L^1$  estandarizado

Tabla 4.15: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 2A. Gráfico basado en el estadístico  $L^2$

Muestra 1	Muestra 2	Método	Potencia						
			0	0.1	0.2	0.3	0.4	0.5	0.6
20	20	Montecarlo	0,038	0,042	0,096	0,208	0,602	0,894	1,000
		Bootstrap	0,048	0,058	0,140	0,316	0,730	0,976	1,000
		Permutaciones	0,038	0,042	0,094	0,236	0,640	0,938	1,000
20	40	Montecarlo	0,028	0,038	0,074	0,248	0,638	0,964	1,000
		Bootstrap	0,028	0,042	0,078	0,276	0,716	0,964	1,000
		Permutaciones	0,042	0,096	0,174	0,558	0,916	0,964	1,000
40	40	Montecarlo	0,054	0,082	0,320	0,842	1,000	1,000	1,000
		Bootstrap	0,062	0,072	0,268	0,800	1,000	1,000	1,000
		Permutaciones	0,044	0,046	0,210	0,734	1,000	1,000	1,000
40	60	Montecarlo	0,048	0,066	0,250	0,812	1,000	1,000	1,000
		Bootstrap	0,048	0,062	0,232	0,782	1,000	1,000	1,000
		Permutaciones	0,036	0,088	0,338	0,926	1,000	1,000	1,000
60	60	Montecarlo	0,048	0,100	0,432	0,994	1,000	1,000	1,000
		Bootstrap	0,050	0,110	0,480	0,998	1,000	1,000	1,000
		Permutaciones	0,036	0,092	0,406	0,992	1,000	1,000	1,000

Gráfico basado en el estadístico  $L^2$  estandarizado

Tabla 4.16: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 2B. Gráfico basado en la distancia de Mahalanobis

Muestra 1	Muestra 2	Método	Potencia						
			0	0.01	0.015	0.02	0.025	0.03	0.035
20	20	Montecarlo	0,042	0,206	0,532	0,828	0,962	0,996	1,000
		Bootstrap	0,044	0,206	0,538	0,836	0,968	0,996	1,000
		Permutaciones	0,038	0,196	0,514	0,820	0,962	0,996	1,000
20	40	Montecarlo	0,040	0,302	0,664	0,924	0,984	1,000	1,000
		Bootstrap	0,032	0,276	0,642	0,910	0,982	1,000	1,000
		Permutaciones	0,054	0,368	0,720	0,940	0,990	1,000	1,000
40	40	Montecarlo	0,046	0,478	0,922	0,994	1,000	1,000	1,000
		Bootstrap	0,054	0,468	0,916	0,994	1,000	1,000	1,000
		Permutaciones	0,054	0,462	0,914	0,994	1,000	1,000	1,000
40	60	Montecarlo	0,046	0,620	0,966	1,000	1,000	1,000	1,000
		Bootstrap	0,048	0,604	0,964	1,000	1,000	1,000	1,000
		Permutaciones	0,048	0,628	0,966	1,000	1,000	1,000	1,000
60	60	Montecarlo	0,054	0,766	0,988	1,000	1,000	1,000	1,000
		Bootstrap	0,054	0,748	0,984	1,000	1,000	1,000	1,000
		Permutaciones	0,052	0,724	0,984	1,000	1,000	1,000	1,000

Gráfico basado en la distancia de Mahalanobis



Tabla 4.17: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 2B. Gráfico basado en el estadístico  $L^1$  estandarizado

Muestra 1	Muestra 2	Método	Potencia						
			0	0.1	0.2	0.3	0.4	0.5	0.6
20	20	Montecarlo	0,044	0,046	0,072	0,174	0,486	0,878	1,000
		Bootstrap	0,058	0,076	0,106	0,284	0,644	0,984	1,000
		Permutaciones	0,050	0,052	0,074	0,206	0,578	0,960	1,000
20	40	Montecarlo	0,044	0,046	0,072	0,174	0,486	0,878	1,000
		Bootstrap	0,058	0,076	0,106	0,284	0,644	0,984	1,000
		Permutaciones	0,050	0,052	0,074	0,206	0,578	0,960	1,000
40	40	Montecarlo	0,054	0,062	0,158	0,616	1,000	1,000	1,000
		Bootstrap	0,064	0,070	0,206	0,782	1,000	1,000	1,000
		Permutaciones	0,054	0,062	0,174	0,700	1,000	1,000	1,000
40	60	Montecarlo	0,048	0,038	0,146	0,616	1,000	1,000	1,000
		Bootstrap	0,042	0,040	0,154	0,728	1,000	1,000	1,000
		Permutaciones	0,052	0,078	0,286	0,930	1,000	1,000	1,000
60	60	Montecarlo	0,050	0,068	0,352	0,996	1,000	1,000	1,000
		Bootstrap	0,054	0,084	0,428	1,000	1,000	1,000	1,000
		Permutaciones	0,050	0,070	0,372	1,000	1,000	1,000	1,000

Gráfico basado en el estadístico  $L^1$  estandarizado

Tabla 4.18: Resultados de la estimación de las tasas de falsas alarmas (error tipo I) y la potencia del gráfico de control  $1 - \beta$  en el escenario 2B. Gráfico basado en el estadístico  $L^2$  estandarizado

Muestra 1	Muestra 2	Método	Potencia						
			0	0.1	0.2	0.3	0.4	0.5	0.6
20	20	Montecarlo	0,044	0,066	0,116	0,356	0,780	0,998	1,000
		Bootstrap	0,056	0,078	0,158	0,472	0,884	1,000	1,000
		Permutaciones	0,044	0,062	0,108	0,356	0,832	1,000	1,000
20	40	Montecarlo	0,032	0,038	0,098	0,310	0,776	0,998	1,000
		Bootstrap	0,026	0,040	0,106	0,384	0,838	0,998	1,000
		Permutaciones	0,060	0,080	0,250	0,672	0,982	1,000	1,000
40	40	Montecarlo	0,046	0,076	0,326	0,946	1,000	1,000	1,000
		Bootstrap	0,046	0,086	0,344	0,968	1,000	1,000	1,000
		Permutaciones	0,044	0,070	0,284	0,930	1,000	1,000	1,000
40	60	Montecarlo	0,044	0,048	0,306	0,946	1,000	1,000	1,000
		Bootstrap	0,054	0,050	0,298	0,946	1,000	1,000	1,000
		Permutaciones	0,048	0,090	0,456	0,992	1,000	1,000	1,000
60	60	Montecarlo	0,050	0,122	0,656	1,000	1,000	1,000	1,000
		Bootstrap	0,048	0,124	0,666	1,000	1,000	1,000	1,000
		Permutaciones	0,056	0,102	0,584	1,000	1,000	1,000	1,000

Gráfico basado en el estadístico  $L^2$  estandarizado

## 4.4 Caso de Estudio: Aneurismas de pacientes de alto y bajo riesgo

El análisis de datos funcionales ha tenido un gran auge en los últimos años, sus aplicaciones abarcan una gran cantidad de áreas desde las ciencias de la salud hasta las finanzas. Algunas veces, en el control estadístico, las variables que van a ser monitorizadas se pueden ver en forma natural como una curva o una función suave. En este sentido, los gráficos de control extendidos para datos funcionales han tomado mayor relevancia.

Para la presente tesis se cuenta con una base de datos de pacientes internados en el Ospedale Niguarda Ca'Granda Milano desde septiembre 2002 a octubre 2005 por sospecha de un aneurisma a lo largo de la arteria carótida interna (ACI). Se dispone de un conjunto de datos referentes a características geométricas y hemodinámicas de los últimos 5cm de la ACI. La base de datos se la encuentra en [48]. Las características mencionadas anteriormente se modelaron en la literatura como datos funcionales (los detalles sobre el preprocesamiento incluido el registro y suavizado se pueden encontrar en [49] y [50]).

La ACI es una arteria principal que controla el suministro de sangre al cerebro, la parte superior de la ACI denominada (ACI-s) se encuentra dentro del cráneo y proporciona sangre al círculo de Willis, mientras la parte inferior (ACI-i) se encuentra fuera del cráneo y extrae sangre de la arteria carótida común. De los pacientes hospitalizados, casi la mitad de ellos tenían un aneurisma en la AIC-s, mientras la otra mitad no tenía aneurisma o tenía aneurisma en la AIC-i. De acuerdo a la localización del aneurisma se clasifican a los pacientes en dos categorías:

1. El grupo de alto riesgo: cuando el aneurisma se aloja dentro del cráneo, por lo general, provoca daños permanentes o letales en el tejido cerebral.
2. El grupo de bajo riesgo: cuando no hay un aneurisma o si el aneurisma está fuera del cráneo, y su posible ruptura no afecta directamente a los tejidos cerebrales.

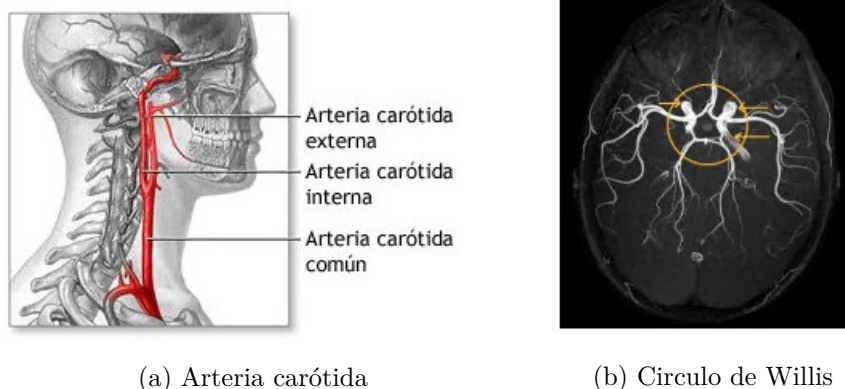


Figura 4.2: Izquierda (a): Imagen de la arteria carótida. Derecha (b): el círculo de Willis, ubicado en la base del cerebro, dentro del cráneo, es una red de pequeñas arterias y capilares que conectan las arterias principales que llevan sangre al cerebro. Fuente: Adaptado por [49]

Además de asociar a cada paciente a uno de los grupos mencionados anteriormente, para un subconjunto de 50 pacientes, el conjunto de datos también tiene una serie de características geométricas y hemodinámicas de los últimos 5cm de la ACI, que se cree que son relevantes para de alguna manera discriminar a los pacientes de alto riesgo de los de bajo riesgo. Eso podría deberse a que los sujetos de alto riesgo tienen un radio mayor en los últimos 5 cm de la ACI. Así también, es factible analizar variación espacial en el radio a lo largo de los últimos 5 cm de la ICA.

Por lo mencionado en el párrafo anterior, para este caso nos enfocaremos en el radio máximo local de una esfera inscrita como función de la longitud del arco y su derivada a lo largo de la línea central de la ACI. En la siguiente figura se ilustra un ejemplo de un vaso reconstruido con un aneurisma considerando el radio máximo de la esfera inscrita.

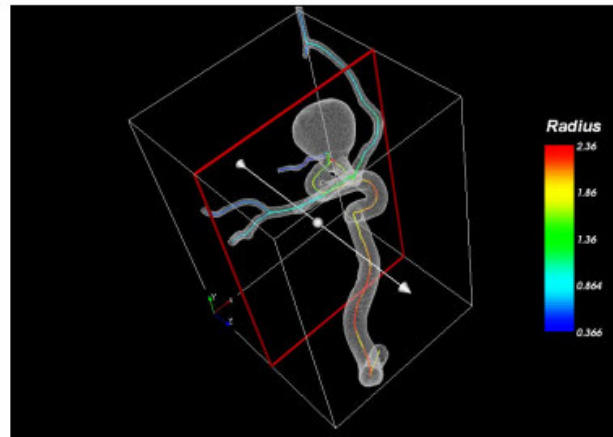


Figura 4.3: Ejemplo de un vaso reconstruido con un aneurisma. La cuadrícula transparente representa la superficie reconstruida, las líneas coloreadas representan las líneas centrales reconstruidas (con el color al radio máximo de la esfera inscrita) y el cubo principal representa la región escaneada

Fuente: Adaptado por [51]

El alcance de esta aplicación será discriminar si el radio del aneurisma o su derivada en los últimos 5 cm de la ACI, discrimina entre los pacientes de alto y bajo riesgo. Además, en ocasiones los pacientes recurren a procedimientos tales como la embolización, que hace que tanto el tamaño y la ondulación del aneurisma varíen, considerando esto, se requeriría dar seguimiento al efecto que ha tenido dicho tratamiento en los pacientes.

A continuación en la siguiente figura se presenta el conjunto de datos del radio y la derivada del radio.

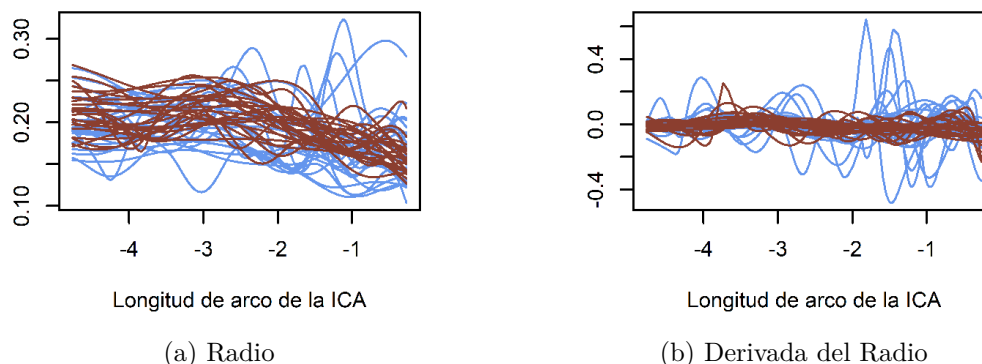


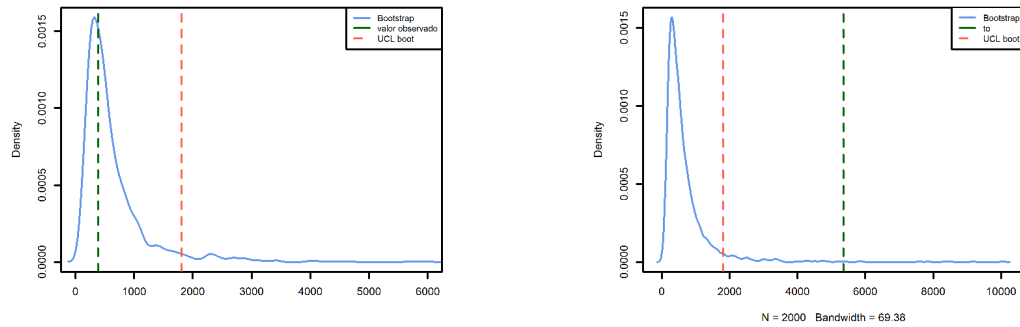
Figura 4.4: Conjunto de datos del radio y derivada del radio: La muestra de curvas de radio (izquierda) y sus primeras derivadas (derecha). Las curvas celestes corresponden al grupo de bajo riesgo y las curvas rojas corresponde del grupo de alto riesgo

Para cumplir con los alcances mencionados anteriormente, primero se ha considerando el radio de la ACI de los pacientes de alto y bajo riesgo se ha realizado el test de igualdad de covarianzas sugerido en por [52], obteniéndose un  $p$  valor de 0.1436, lo que determina que no se rechaza la hipótesis nula, es decir, que tanto los pacientes de alto y bajo riesgo presentan igual estructura de la covarianza. Por otra parte, al analizar la derivada del radio se tiene un resultado similar ya que el  $p$  valor es igual a 0.3096, por ende, se no se rechaza la hipótesis nula.

Adicionalmente, se ha realizado un test de estacionariedad propuesto en [53]. Al evaluar el radio de los pacientes de bajo y alto riesgo, se han obtenido los  $p$  valores de 0.0318 y 0.0004 respectivamente, lo que significa que a un nivel de significancia del 5% se rechaza la hipótesis nula de estacionariedad. Por otra parte, al analizar la derivada del radio los valores  $p$  fueron de 0.0116 y 0.0106, por ende, también se rechaza la hipótesis nula de estacionariedad. De acuerdo a lo mencionado y al estudio de simulación se ha optado por aplicar el estadístico basado en la distancia de Mahalanobis a partir del procedimiento bootstrap para diseñar la prueba de hipótesis y el respectivo gráfico de control para dar seguimiento al tratamiento.

Se han realizado las pruebas de hipótesis iniciales del gráfico de control, donde se

contrasta la media de los pacientes de alto y bajo riesgo. En la siguiente figura se ilustran dichas pruebas. Se observa que el radio del aneurisma no permite discriminar los pacientes, sin embargo, la derivada del aneurisma permite establecer que la media de los pacientes de alto y bajo riesgo son estadísticamente diferentes.



(a) Prueba de hipótesis considerando la distancia de Mahalanobis para el radio del aneurisma del grupo de pacientes de alto y bajo riesgo (b) Prueba de hipótesis considerando la distancia de Mahalanobis para la derivada del radio para el grupo de pacientes de alto y bajo riesgo

Figura 4.5

Ahora procederemos a evaluar el gráfico de control en el tiempo, considerando que los pacientes evaluados en el tiempo han recibido determinado tratamiento que modifica la el radio y su ondulación. Este gráfico de control nos permite establecer la eficacia de dicho tratamiento en el tiempo, es decir, después de qué tiempo los pacientes de bajo riesgo vuelven a un estado alto de riesgo.

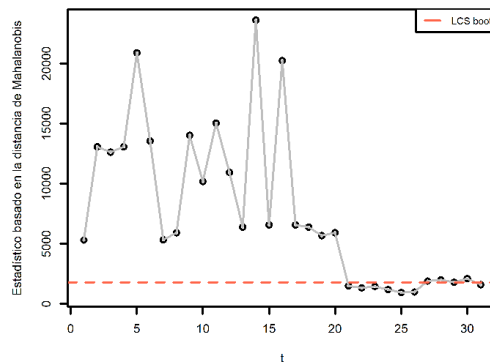


Figura 4.6: Gráfico de control para la derivada del radio del grupo de pacientes de alto y bajo riesgo

## 5.1 Conclusiones y líneas futuras

### 5.1.1 Conclusiones

Se ha propuesto una nueva alternativa para el diseño del gráfico de control cuando existen múltiples variables críticas de calidad de un proceso funcional. Esta propuesta se basa en el diseño de la monitorización del gráfico de control basado en la distancia de Mahalanobis,  $L^1$  estandarizado y  $L^2$  estandarizado. Para este desarrollo se han propuesto inicialmente el uso de métodos de detección de atípicos para datos funcionales presentado por Febrero-Bande et al. en [32], y para encontrar el límite de control superior de la distribución del estadístico se han usado técnicas de remuestreo bootstrap suavizado y permutaciones recomendado en [27].

Se ha realizado un estudio de simulación que permite comparar el desempeño (medido como porcentaje de rechazo cuando no se cumple la hipótesis nula) dependiendo del estadístico usado (distancia de Mahalanobis,  $L^1$  estandarizado y  $L^2$  estandarizado) y del procedimiento para determinar el límite de control superior (bootstrap suavizado o permutaciones).

En el estudio de simulaciones se diseñaron escenarios para la detección de cambios de la media en desplazamiento y forma considerando operadores de covarianza estacionario y no estacionario. Se tuvo un desempeño superior del gráfico de control basado en la distancia de Mahalanobis cuando se requiere detectar cambios en el desplazamiento de la media para el operador de covarianza no estacionario, teniendo un mejor desempeño para el método de bootstrap a medida que se incrementa el



tamaño de la muestra de monitoreo y de control. Por otra parte, cuando es necesario detectar cambios en desplazamiento de media y las trayectorias son estacionarias tienen un mejor desempeño los gráficos de control  $L^1$  y  $L^2$  estandarizadas. Así también, si se quiere detectar cambios de la media de un proceso en forma, el gráfico de control basado en la distancia de Mahalanobis tiene mayor potencia para el operador de covarianza estacionario y no estacionario.

Esta propuesta metodológica ha sido aplicada con éxito al caso de monitoreo de la derivada del radio a lo largo de la arteria carótida interna de los pacientes de alto y bajo riesgo. Se determinó que un buen biomarcador para discriminar a los pacientes de alto y bajo riesgo es la derivada del radio del aneurisma a lo largo de la arteria carótida interna, lo cual indica el impacto de la elección del espacio de Hilbert (curvas suaves  $L^2$  o sus derivadas) para el análisis de los datos. El gráfico de control formulado es útil para dar un seguimiento al procedimiento optado por los pacientes.

### 5.1.2 Líneas Futuras

Las líneas futuras de investigación en el campo de gráficos de control están relacionadas con el diseño de gráficos de control Shewhart de la media, donde se consideren diferentes estadísticos que tomen en cuenta autocorrelación y estructuras de covarianza y varianza, considerando las muestras subagrupadas y observaciones individuales. Se requiere una amplia labor investigativa para el diseño de gráficos de control de perfiles que tomen en cuenta técnicas del análisis de datos funcional, en vista de que en el campo de la estadística, la recopilación de información es en tiempo real y/o con referencias espaciales lo que produce datos masivos de alta dimensión que es recomendable analizarlos desde un enfoque funcional.

Considerando también que los gráficos de control pueden ser vistos como pruebas de hipótesis, en la actualidad existe un gran desafío al momento de realizar inferencia estadística para datos funcionales, ya que es un caso extremo de datos de alta dimensión, para los que independientemente del tamaño de la muestra, la información tiende a ser insuficiente para caracterizar completamente un modelo subyacente, por

tanto, se requiere de una alta labor investigativa para que los estadísticos tradicionales que se han venido usando sean aptos para abordar problemas de alta dimensionalidad.

En este trabajo se ha desarrollado la metodología para el diseño de la monitorización del gráfico de control basado en la distancia de Mahalanobis,  $L^1$  estandarizado y  $L^2$  estandarizado considerando técnicas del análisis de datos funcionales como bootstrap suavizado, permutaciones y medidas de profundidad; sin embargo, de acuerdo a las necesidades del usuario aún existe la oportunidad de variar dichas cartas de control dependiendo si se desear monitorear los datos individualmente (como lo es en la fase de detección de atípicos) o datos subagrupados.

```
#####  
#####Escenario 1A#####  
  
#Parámetros escenario 1A  
  
mc<-1000  
n0<-25  
n1<-25  
n<-n1+n0  
idx1 <- seq_len(n0)  
idx2 <- seq_len(n0) + n1  
tt <-seq( 0, 1, 0.2)  
var.teor<- 1  
mu0<- 30* tt * (1 - tt)^(3/2)  
corr.teor<-outer(tt,tt,function(u,v){exp(-2*(u-v)^2)})  
rho <- 0  
m_h0<-vector(length = mc)  
d1_h0<-list()  
d2_h0<-list()  
  
#Función para simular data  
  
func.sim.set <- function(t, var.teor=1, trend.teor, corr.teor , rho = 0)  
{  
  mdata <- length(t)  
  sd.teor <- sqrt(as.numeric(var.teor))  
  if (rho != 0) corr.teor <- corr.teor * sqrt((1+rho)/(1-rho))  
  C<-svd(t(corr.teor))  
  L.corr.teor=(C$u%*%diag(sqrt(C$d))  
  # L.corr.teor <- t(chol(corr.teor))
```

```

# Generacion de datos simulados:  $Y = m(X) + s(X) * E$ 
func.sim <- function(rep){
  err.norm <- matrix(rnorm(mdata * rep), nrow = mdata)
  data.err <- L.corr.teor %*% err.norm
  if (rho != 0) {
    data.err[, 1] <- data.err[, 1] * sqrt((1-rho)/(1+rho))
    for(i in 2:rep)
      data.err[, i] <- rho*data.err[, i-1] + (1 - rho)*data.err[, i]
  }

  res <- as.numeric(trend.teor) + data.err
  return(res)
}
return(func.sim)
}

f0 <- func.sim.set(tt, var.teor, trend.teor=mu0, corr.teor=corr.teor, rho)

#Montecarlo bajo h0

for(i in 1:mc)
{
  set.seed(i)
  df<-fdata(t(f0(n0+n1)))$data
  d1_h0[[i]]<- df[idx1, , drop = FALSE]
  d2_h0[[i]]<- df[idx2, , drop = FALSE]
  m_h0[i]<-stat_hotelling(d1_h0[[i]],d2_h0[[i]],paired = FALSE,step_size = 0.5)
}

UCL_mc<-quantile(m_h0,1-0.05)

save.image("Escenario_1_A/t2_25_25_delta_est/RData/mc_h0_25_25_delta0_est
.RData")

load("Escenario_1_A/t2_25_25_delta_est/RData/mc_h0_25_25_delta0_est.RData")

t2_ha<-vector(length = mc)

```

```

d1_ha<-list()
delta<-0.2

f1 <- func.sim.set(tt, var.teor, mu0+delta, corr.teor, rho = rho)

for(i in 1:mc)
{
d1_ha[[i]]<-fdata(t(f1(n1)))$data
t2_ha[i]<-stat.hotelling(x=d1_h0[[i]],y=d1_ha[[i]],paired=FALSE,step_size = 0.5)
}

save.image("Escenario_1_A/t2_25_25_delta_est/RData/mc_ha_25_25_delta0.2_est
.RData")

load("Escenario_1_A/t2_25_25_delta_est/RData/mc_h0_25_25_delta0_est.RData")
load("Escenario_1_A/t2_25_25_delta_est/RData/mc_ha_25_25_delta0.2_est.RData")

#Parámetros escenario 1A bootstrap

boot<-500
t2_h0<-matrix(0,nrow=mc,ncol=boot)
UCL_boot<-vector(length = mc)
data_trim<-list()

trim.data<-function(d_h0,d_ha)
{
data<-rbind(d_h0,d_ha)
out.FM<-depth.FM(data,trim=0.10,draw=TRUE,scale=T)
data_trim<-data[out.FM$ltrim,]
return(data_trim)
}

func.sim.boot<-function(data_trim)
{
t2_boot_h0<-vector(length= boot)
cov.est=var(data_trim)

```

```

r <- eigen(cov.est, symmetric = TRUE)
v <- r$vectors
lambda <- r$values
smo <- 0.09
tol <- 1e-06
if (!all(lambda >= -tol * abs(lambda[1L])))
  stop("'Sigma' is not positive definite")
L.boot <- v %%% diag(sqrt(pmax(lambda*smo, 0)), ncol(data_trim))

func.boot<- function(boot)
{
  for(j in 1:boot)
  {
    ind<-sample(1:nrow(data_trim),n,replace = TRUE)
    dat<-data_trim[ind,]+t(L.boot %%% matrix(rnorm(n* ncol(data_trim)),
                                             nrow = ncol(data_trim)))

    datboot1<-dat[idx1,]
    datboot2<-dat[idx2,]
    t2_boot_h0[j]<-stat_hotelling(datboot1,datboot2,step_size = 0.01,
                                paired = FALSE)
  }
  return(t2_boot_h0)
}
return(func.boot)
}

for(i in 1:mc)
{
  data_trim[[i]]<-trim.data(d1_h0[[i]],d1_ha[[i]])
  boot_data_trim<-func.sim.boot(data_trim[[i]])
  t2_h0[i,]<-boot_data_trim(boot)
  UCL_boot[i]<-quantile(t2_h0[i,],1-0.05)
}

save.image("Escenario_1_A/t2_25_25_delta_est/RData/boot_25_25_delta0.2_est
.RData")

load("Escenario_1_A/t2_25_25_delta_est/RData/mc_h0_25_25_delta0_est.RData")

```

```

load("Escenario_1_A/t2_25_25_delta_est/RData/mc_ha_25_25_delta0.2_est.RData")
load("Escenario_1_A/t2_25_25_delta_est/RData/boot_25_25_delta0.2_est.RData")

#Permulaciones escenario 1A
t2_per_h0<-vector(length = mc )
px <- n0 / (n0 + n1)
py <- n0 / (n0 + n1)

#Funciones extraídas de Fda Hotelling
partition <- function(df, n, probs) {
  if (is.null(dim(n))) {
    return(n %>%
      replicate(
        expr = split_matrix(df, random_group(nrow(df), probs)),
        simplify = FALSE
      ) %>%
      purrr::transpose() %>%
      dplyr::as_data_frame()
    )
  }
  dplyr::data_frame(combId = seq_len(dim(n)[2])) %>%
    dplyr::mutate(D1 = purrr::map(combId, ~ df[ n[, .x], ]),
                 D2 = purrr::map(combId, ~ df[-n[, .x], ])) %>%
    dplyr::select(-combId) %>%
    purrr::set_names(names(probs))
}

random_group <- function(n, probs) {
  probs <- probs / sum(probs)
  g <- findInterval(seq(0, 1, length = n), c(0, cumsum(probs)),
                    rightmost.closed = TRUE)
  names(probs)[sample(g)]
}

split_matrix <- function(x, indices) {
  nn <- levels(as.factor(indices))

```

```

res <- list(x[indices == nn[1], , drop = FALSE],
           x[indices == nn[2], , drop = FALSE])
names(res) <- nn
res
}

for(i in 1:mc)
{
  stat_perm<- data_trim[[i]] %>%partition(mc, c(D1 = px, D2 = py))
  %>%dplyr::mutate(stat_boot = purrr::map2(D1, D2, stat_hotelling,
    mu = NULL, paired = FALSE, step_size = 0.05))
  t2_per_h0[i]<-unlist(stat_perm$stat_boot)
}

UCL_per<-quantile(t2_per_h0,1-0.05)

save.image("Escenario_1_A/t2_25_25_delta_est/RData/per_25_25_
          delta0.2_est.RData")

#####
#####Escenario 1B#####

#Parámetros escenario 1B
mc<-1000
n0<-25
n1<-25
n<-n1+n0
idx1 <- seq_len(n0)
idx2 <- seq_len(n0) + n1
tt <-seq( 0, 1, 0.2)
var.teor<- 1
mu0<- 30* tt * (1 - tt)^(3/2)
corr.teor<-outer(tt,tt,function(u,v){exp(-2*(u-v)^2)*(u+0.5)*(v+0.5)})
rho <- 0
m_h0<-vector(length = mc)
d1_h0<-list()
d2_h0<-list()

```



```
f0 <- func.sim.set(tt, var.teor, trend.teor=mu0, corr.teor=corr.teor, rho)

for(i in 1:mc)
{
  set.seed(i)
  df<-fdata(t(f0(n0+n1)))$data
  d1_h0[[i]]<- df[idx1, , drop = FALSE]
  d2_h0[[i]]<- df[idx2, , drop = FALSE]
  m_h0[i]<-stat.hotelling(d1_h0[[i]],d2_h0[[i]],paired = FALSE,step_size = 0.5)
}

UCL_mc<-quantile(m_h0,1-0.05)

save.image("Escenario_1_B/t2_25_25_delta_no_est/RData/mc_h0_25_25_delta0_no_
          est.RData")

load("Escenario_1_B/t2_25_25_delta_no_est/RData/mc_h0_25_25_delta0
      _no_est.RData")

t2_ha<-vector(length = mc)
d1_ha<-list()
delta<-0.2

f1 <- func.sim.set(tt, var.teor, mu0+delta, corr.teor, rho = rho)

for(i in 1:mc)
{
  d1_ha[[i]]<-fdata(t(f1(n1)))$data
  t2_ha[i]<-stat.hotelling(x=d1_h0[[i]],y=d1_ha[[i]],paired=FALSE,step_size = 0.5)
}

save.image("Escenario_1_B/t2_25_25_delta_no_est/RData/mc_ha_25_25_delta0.2
          _no_est.RData")

load("Escenario_1_B/t2_25_25_delta_no_est/RData/mc_h0_25_25_delta0
      _no_est.RData")

load("Escenario_1_B/t2_25_25_delta_no_est/RData/mc_ha_25_25_delta0.2
```

```

    _no_est.RData")

#Parámetros bootstrap
boot<-500
t2_h0<-matrix(0,nrow=mc,ncol=boot)
UCL_boot<-vector(length = mc)
data_trim<-list()

trim.data<-function(d_h0,d_ha)
{
  data<-rbind(d_h0,d_ha)
  out.FM<-depth.FM(data,trim=0.10,draw=TRUE,scale=T)
  data_trim<-data[out.FM$ltrim,]
  return(data_trim)
}

func.sim.boot<-function(data_trim)
{
  t2_boot_h0<-vector(length= boot)
  cov.est=var(data_trim)
  r <- eigen(cov.est, symmetric = TRUE)
  v <- r$vector
  lambda <- r$values
  smo <- 0.09
  # cov.boot <- cov.est*smo
  tol <- 1e-06
  if (!all(lambda >= -tol * abs(lambda[1L])))
    stop("'Sigma' is not positive definite")
  L.boot <- v %*% diag(sqrt(pmax(lambda*smo, 0)), ncol(data_trim))

func.boot<- function(boot)
{
  for(j in 1:boot)
  {
    ind<-sample(1:nrow(data_trim),n,replace = TRUE)
    dat<-data_trim[ind,]+t(L.boot %*% matrix(rnorm(n* ncol(data_trim)),

```

```

                                nrow = ncol(data_trim)))

datboot1<-dat[idx1,]
datboot2<-dat[idx2,]
t2_boot_h0[j]<-stat_hotelling(datboot1,datboot2,step_size = 0.01,
                             paired = FALSE)
}
return(t2_boot_h0)
}
return(func.boot)
}

for(i in 1:mc)
{
data_trim[[i]]<-trim.data(d1_h0[[i]],d1_ha[[i]])
boot_data_trim<-func.sim.boot(data_trim[[i]])
t2_h0[i,]<-boot_data_trim(boot)
UCL_boot[i]<-quantile(t2_h0[i,],1-0.05)
}

save.image("Escenario_1_B/t2_25_25_delta_no_est/RData/boot_25_25_delta0.2
           _no_est.RData")

load("Escenario_1_B/t2_25_25_delta_no_est/RData/mc_h0_25_25_delta0
     _no_est.RData")
load("Escenario_1_B/t2_25_25_delta_no_est/RData/mc_ha_25_25_delta0.2
     _no_est.RData")
load("Escenario_1_B/t2_25_25_delta_no_est/RData/boot_25_25_delta0.2
     _no_est.RData")

#Permutaciones
t2_per_h0<-vector(length = mc )
px <- n0 / (n0 + n1)
py <- n0 / (n0 + n1)

for(i in 1:mc) #i=1
{

```

```

stat_perm<- data_trim[[i]] %>%partition(mc, c(D1 = px, D2 = py))%>%
  dplyr::mutate(stat_boot = purrr::map2(D1, D2, stat_hotelling,
                                       mu = NULL, paired = FALSE,
                                       step_size = 0.05))
t2_per_h0[i]<-unlist(stat_perm$stat_boot)
}

UCL_per<-quantile(t2_per_h0,1-0.05)

save.image("Escenario_1_B/t2_25_25_delta_no_est/RData/per_25_25_delta0.2_
          no_est.RData")

#####
#####Escenario 2A#####

#Parámetros
mc<-1000
n0<-25
n1<-25
n<-n1+n0
idx1 <- seq_len(n0)
idx2 <- seq_len(n0) + n1
tt <-seq( 0, 1, 0.2)
var.teor<- 1
eta<-0
mu0<-(1-eta)*30*tt*(1-tt)^(3/2)+eta*30*tt^(3/2)*(1-tt)
corr.teor<-outer(tt,tt,function(u,v){exp(-2*(u-v)^2)})
rho <- 0
m_h0<-vector(length = mc)
d1_h0<-list()
d2_h0<-list()

f0 <- func.sim.set(tt, var.teor, trend.teor=mu0, corr.teor=corr.teor, rho)

for(i in 1:mc)
{
  set.seed(i)
  df<-fdata(t(f0(n0+n1)))$data

```

```

d1_h0[[i]]<- df[idx1, , drop = FALSE]
d2_h0[[i]]<- df[idx2, , drop = FALSE]
m_h0[i]<-stat_hotelling(d1_h0[[i]],d2_h0[[i]],paired = FALSE,step_size = 0.5)
}

UCL_mc<-quantile(m_h0,1-0.05)

save.image("Escenario_2A/t2_25_25_delta_no_est/RData/mc_h0_25_25_delta0_no_
_est.RData")

load("Escenario_2_A/t2_25_25_delta_no_est/RData/mc_h0_25_25_delta0_no_
_est.RData")

t2_ha<-vector(length = mc)
d1_ha<-list()
eta<-0.2
mu1<-(1-eta)*30*tt*(1-tt)^(3/2)+eta*30*tt^(3/2)*(1-tt)
f1 <- func.sim.set(tt, var.teor, mu1, corr.teor, rho = rho)

for(i in 1:mc)
{
d1_ha[[i]]<-fdata(t(f1(n1)))$data
t2_ha[i]<-stat_hotelling(x=d1_h0[[i]],y=d1_ha[[i]],paired=FALSE,step_size = 0.5)
}

save.image("Escenario_2_A/t2_25_25_delta_no_est/RData/mc_ha_25_25_delta0.2_
no_est.RData")

load("Escenario_2_A/t2_25_25_delta_no_est/RData/mc_h0_25_25_delta0_no_est.RData")
load("Escenario_2_A/t2_25_25_delta_no_est/RData/mc_ha_25_25_delta0.2_
no_est.RData")

#Parametros Bootstrap

boot<-500

```

```

t2_h0<-matrix(0,nrow=mc,ncol=boot)
UCL_boot<-vector(length = mc)
data_trim<-list()

trim.data<-function(d_h0,d_ha)
{
  data<-rbind(d_h0,d_ha)
  out.FM<-depth.FM(data,trim=0.10,draw=TRUE,scale=T)
  data_trim<-data[out.FM$ltrim,]
  return(data_trim)
}

func.sim.boot<-function(data_trim)
{
  t2_boot_h0<-vector(length= boot)
  cov.est=var(data_trim)
  r <- eigen(cov.est, symmetric = TRUE)
  v <- r$vector
  lambda <- r$values
  smo <- 0.09
  tol <- 1e-06
  if (!all(lambda >= -tol * abs(lambda[1L])))
    stop("'Sigma' is not positive definite")
  L.boot <- v %*% diag(sqrt(pmax(lambda*smo, 0)), ncol(data_trim))

func.boot<- function(boot)
{
  for(j in 1:boot)
  {
    ind<-sample(1:nrow(data_trim),n,replace = TRUE)
    dat<-data_trim[ind,]+t(L.boot %*% matrix(rnorm(n* ncol(data_trim)), nrow = ncol(data_trim)))
    datboot1<-dat[idx1,]
    datboot2<-dat[idx2,]
    t2_boot_h0[j]<-stat_hotelling(datboot1,datboot2,step_size = 0.01,paired = FALSE)
  }
  return(t2_boot_h0)
}

```

```

return(func.boot)
}

for(i in 1:mc)
{
data_trim[[i]]<-trim.data(d1_h0[[i]],d1_ha[[i]])
boot_data_trim<-func.sim.boot(data_trim[[i]])
t2_h0[i,]<-boot_data_trim(boot)
UCL_boot[i]<-quantile(t2_h0[i,],1-0.05)
}

save.image("Escenario_2_A/t2_25_25_delta_no_est/RData/boot_25_25_delta0.2_
          no_est.RData")

load("Escenario_2_A/t2_25_25_delta_no_est/RData/mc_h0_25_25_delta0_
      no_est.RData")
load("Escenario_2_A/t2_25_25_delta_no_est/RData/mc_ha_25_25_delta0.2_
      _no_est.RData")
load("Escenario_2_A/t2_25_25_delta_no_est/RData/boot_25_25_delta0.2_
      _no_est.RData")

#Permulaciones escenario 2A
t2_per_h0<-vector(length = mc )
px <- n0 / (n0 + n1)
py <- n0 / (n0 + n1)

for(i in 1:mc)
{
stat_perm<- data_trim[[i]] %>%partition(mc, c(D1 = px, D2 = py))
%>%dplyr::mutate(stat_boot = purrr::map2(D1, D2, stat_hotelling, mu = NULL,
                                       paired = FALSE,step_size = 0.05))
t2_per_h0[i]<-unlist(stat_perm$stat_boot)
}

UCL_per<-quantile(t2_per_h0,1-0.05)

```

```

save.image("Escenario_2_B/t2_25_25_delta_no_est/RData/per_25_25_delta0.2
           _no_est.RData")

#####
#####Escenario 2B#####

#Parámetros escenario 2B

mc<- 1000
n0<-25
n1<-25
n<-n1+n0
idx1 <- seq_len(n0)
idx2 <- seq_len(n0) + n1
tt <-seq( 0, 1, 0.2)
var.teor<- 1
eta<-0
mu0<-(1-eta)*30*tt*(1-tt)^(3/2)+eta*30*tt^(3/2)*(1-tt)
corr.teor<-outer(tt,tt,function(u,v){exp(-2*(u-v)^2)*(u+0.5)*(v+0.5)})
rho <- 0
m_h0<-vector(length = mc)
d1_h0<-list()
d2_h0<-list()

f0 <- func.sim.set(tt, var.teor, trend.teor=mu0, corr.teor=corr.teor, rho)

for(i in 1:mc)
{
  set.seed(i)
  df<-fdata(t(f0(n0+n1)))$data
  d1_h0[[i]]<- df[idx1, , drop = FALSE]
  d2_h0[[i]]<- df[idx2, , drop = FALSE]
  m_h0[i]<-stat_hotelling(d1_h0[[i]],d2_h0[[i]],paired = FALSE,step_size = 0.5)
}

UCL_mc<-quantile(m_h0,1-0.05)

save.image("Escenario_2_B/t2_25_25_delta_no_est/RData/mc_h0_25_25_delta0

```



```

        _no_est.RData")

load("Escenario_2_B/t2_25_25_delta_no_est/RData/mc_h0_25_25_delta0
      _no_est.RData")

t2_ha<-vector(length = mc)
d1_ha<-list()
eta<-0.2
mu1<-(1-eta)*30*tt*(1-tt)^(3/2)+eta*30*tt^(3/2)*(1-tt)
f1 <- func.sim.set(tt, var.teor, mu1, corr.teor, rho = rho)

for(i in 1:mc)
{
d1_ha[[i]]<-fdata(t(f1(n1)))$data
t2_ha[i]<-stat.hotelling(x=d1_h0[[i]],y=d1_ha[[i]],paired=FALSE,step_size = 0.5)
}

save.image("Escenario_2_B/t2_25_25_delta_no_est/RData/
           mc_ha_25_25_delta0.2_no_est.RData")

load("Escenario_2_B/t2_25_25_delta_no_est/RData/mc_h0_25_25_delta0_no_est.RData")
load("Escenario_2_B/t2_25_25_delta_no_est/RData/mc_ha_25_25_delta0.2
      _no_est.RData")

boot<-500
t2_h0<-matrix(0,nrow=mc,ncol=boot)
UCL_boot<-vector(length = mc)
data_trim<-list()

for(i in 1:mc)
{
data_trim[[i]]<-trim.data(d1_h0[[i]],d1_ha[[i]])
boot_data_trim<-func.sim.boot(data_trim[[i]])

```

```
t2_h0[i,]<-boot_data_trim(boot)
UCL_boot[i]<-quantile(t2_h0[i,],1-0.05)
}

save.image("Escenario_2_B/t2_25_25_delta_no_est/RData/boot_25_25_delta0.2_
          no_est.RData")

load("Escenario_2_B/t2_25_25_delta_no_est/RData/mc_h0_25_25_delta0
      _no_est.RData")
load("Escenario_2_B/t2_25_25_delta_no_est/RData/mc_ha_25_25_delta0.2
      _no_est.RData")
load("Escenario_2_B/t2_25_25_delta_no_est/RData/boot_25_25_delta0.2
      _no_est.RData")

t2_per_h0<-vector(length = mc )
px <- n0 / (n0 + n1)
py <- n0 / (n0 + n1)

for(i in 1:mc)
{
  stat_perm<- data_trim[[i]] %>%partition(mc, c(D1 = px, D2 = py))
  %>%dplyr::mutate(stat_boot = purrr::map2(D1, D2, stat_hotelling, mu = NULL,
                                          paired = FALSE,step_size = 0.05))
  t2_per_h0[i]<-unlist(stat_perm$stat_boot)
}

UCL_per<-quantile(t2_per_h0,1-0.05)

save.image("Escenario_2_B/t2_25_25_delta_no_est/RData/per_25_25_delta0.2
          no_est.RData")
```

Una de las características distintivas de los espacios euclídeos  $p$  – dimensionales es la facilidad con la que se puede definir métricas, las mismas que satisfacen las siguientes propiedades.

**Definición B.0.1.** *Se define una métrica sobre un conjunto  $\mathbb{M}$  como una función  $d : \mathbb{M} \times \mathbb{M} \rightarrow \mathbb{R}$  que satisface*

1.  $d(x_1, x_2) \geq 0$ ,
2.  $d(x_1, x_2) = 0$  si  $x_1 = x_2$
3.  $d(x_1, x_3) \leq d(x_1, x_2) + d(x_2, x_3)$

para todo  $x_1, x_2, x_3 \in \mathbb{M}$ . Llamamos espacio métrico a la dupla  $(\mathbb{M}, d)$ .

**Definición B.0.2.** *Sea  $(\mathbb{M}, d)$  un espacio métrico con  $E \subset \mathbb{M}$ . Entonces, se dice que  $E$  es abierto si para todo  $e \in E$  existe un  $\epsilon > 0$  tal que  $\{x \in \mathbb{M} : d(x, e) < \epsilon\}$ . El subconjunto  $E$  es cerrado si  $\{x \in \mathbb{M} : x \notin E\}$  es abierto.*

A continuación se presentan otros conceptos clásicos relacionados con el espacio métrico.

**Definición B.0.3.** *La clausura  $\overline{E}$  de  $E \subset \mathbb{M}$  es el conjunto cerrado más pequeño de  $\mathbb{M}$  que contiene a  $E$ .*

**Definición B.0.4.** *Un conjunto  $E \subset \mathbb{M}$  es denso en  $\mathbb{M}$  si  $\overline{E} = \mathbb{M}$ .*

**Definición B.0.5.** *Un espacio métrico  $\mathbb{M}$  es separable si tiene un subconjunto denso y contable.*

Si tenemos una función de un espacio métrico a otro entonces su suavidad (es decir, cómo cambia el valor de la función a medida que el argumento cambia) es una consideración importante para los datos funcionales. La forma más básica de la suavidad es la continuidad.

**Definición B.0.6.** Sea  $(\mathbb{M}_1, d_1)$  y  $(\mathbb{M}_2, d_2)$  espacios métricos y  $f$  una función definida sobre  $\mathbb{M}_1$  que toma valores en  $\mathbb{M}_2$ . Entonces  $f$  es continua en  $x \in \mathbb{M}_1$  si para todo  $\epsilon > 0$  existe  $\delta_{x,\epsilon} > 0$  tal que para todo  $y \in \mathbb{M}_1$ ,  $d_1(x, y) < \delta_{x,\epsilon}$  implica que  $d_2(f(x), f(y)) < \epsilon$ . La función  $f$  es uniformemente continua cuando  $\delta_{x,\epsilon}$  no depende de  $x$ .

**Definición B.0.7.** Sea  $(\mathbb{M}, d)$  un espacio métrico con  $\{x_n\}$  una sucesión de puntos en  $\mathbb{M}$ . La sucesión converge a  $x \in \mathbb{M}$ , denotado por  $x_n \rightarrow x$ , si  $d(x_n, x) \rightarrow 0$  cuando  $n \rightarrow \infty$ .

Con la notación anterior, decimos que  $f$  es continua en  $x$  si  $f(x_n) \rightarrow f(x)$  cuando  $x_n \rightarrow x$ .

**Definición B.0.8.** Una sucesión de elementos  $\{x_n\}$  en un espacio métrico  $(\mathbb{M}, m)$  es una sucesión de Cauchy si  $\sup_{m,n \geq N} d(x_m, x_n) \rightarrow 0$  cuando  $N \rightarrow \infty$ .

**Definición B.0.9.** Un espacio métrico  $(\mathbb{M}, m)$  es completo si toda sucesión de Cauchy es convergente.

**Definición B.0.10.** Un espacio métrico  $\mathbb{M}$  es compacto si toda sucesión de elementos en  $\mathbb{M}$  tiene una subsucesión que converge a un elemento de  $\mathbb{M}$ .

**Definición B.0.11.** Un espacio métrico  $(\mathbb{M}, d)$  es acotado totalmente si para todo  $\epsilon > 0$  existe  $y_1, \dots, y_n \in \mathbb{M}$  tal que  $\mathbb{M} \subset \cup_{i=1}^n \{x \in \mathbb{M} : d(x, y_i) < \epsilon\}$  para algún entero positivo  $n$ .

El resultado anterior es conocido como el teorema Heine-Borel.

**Teorema B.0.1.** Un espacio métrico es compacto si es completo y totalmente acotado.

*Demostración.* La demostración véase en [33, p. 123].

## B.0.1 Espacios vectoriales y normados

Los espacios métricos representan una abstracción topológica natural de  $\mathbb{R}^p$  que es adecuada para nuestros propósitos; sin embargo, se requiere de una estructura algebraica, para esto, en esta sección se introduce el concepto de espacio vectorial.

**Definición B.0.12.** *Un espacio vectorial  $\mathbb{V}$  es un conjunto de elementos llamados vectores. Sean  $v_1, v_2$  vectores la suma retorna otro vector, denotado por  $v_1 + v_2$ . Dado un vector  $v$  y  $a \in \mathbb{R}$ , la multiplicación de un vector por un escalar retorna un vector denotado por  $av$ . La suma y la multiplicación satisfacen las siguientes propiedades.*

1.  $v_1 + v_2 = v_2 + v_1$
2.  $v_1 + (v_2 + v_3) = (v_1 + v_2) + v_3$
3.  $a_1(a_2v) = (a_1a_2)v$
4.  $a(v_1 + v_2) = av_1 + av_2$
5.  $(a_1 + a_2)v = a_1v + a_2v$
6.  $1v = v$

*Además, existe un único elemento  $0$  tal que  $v + 0 = v$  para todo  $v \in \mathbb{V}$ , y a cada elemento  $v$  corresponde un elemento  $-v$  tal que  $v + (-v) = 0$ .*

A menudo trabajaremos con subconjuntos de espacios vectoriales. Cuando un subconjunto admite las mismas operaciones algebraicas que su espacio vectorial, lo llamamos un subespacio lineal o simplemente como un subespacio. En particular, comenzando con un subconjunto  $A$  de un espacio vectorial  $\mathbb{V}$ , podemos definir un subespacio formando un nuevo conjunto que contiene todas las combinaciones lineales de dimensión finita de sus elementos. Este se denomina *span* de  $A$  y se denota como  $\text{span}(A)$ . Una caracterización del  $\text{span}(A)$  es como la intersección de todos los subespacios en  $\mathbb{V}$  que contienen a  $A$ . En este sentido,  $\text{span}(A)$  es el subespacio más pequeño que contiene los elementos de  $A$ .

**Definición B.0.13.** Sea  $\mathbb{V}$  un espacio vectorial y  $B = \{v_1, \dots, v_k\} \subset \mathbb{V}$  para algún entero finito y positivo  $k$ . La colección  $B$  es linealmente independiente si  $\sum_{i=1}^k a_i v_i = 0$  implica  $a_i = 0$  para todo  $i$ .

**Definición B.0.14.** Si  $B = \{v_1, \dots, v_k\}$  es un subconjunto linealmente independiente del espacio vectorial  $\mathbb{V}$  y el  $\text{span}(B) = \mathbb{V}$ , se dice que  $B$  es una base de  $\mathbb{V}$ .

**Teorema B.0.2.** Si  $\mathbb{V}$  es un espacio vectorial con norma  $\|\cdot\|$ , entonces se sigue que  $d(x, y) := \|x - y\|$  para todo  $x, y \in \mathbb{V}$  es un métrica.

*Demostración.* La demostración véase en [33, p. 200].

**Definición B.0.15.** Sea  $A$  un subconjunto de un espacio vectorial normado, el conjunto cerrado del  $\text{span}$  de  $A$ , denotado por  $\overline{\text{span}(A)}$  es la clausura del  $\text{span}(A)$  con respecto a métrica inducida por la norma.

## B.0.2 Espacios de Hilbert y dotados de un producto inercial

**Definición B.0.16.** Una función  $\langle \cdot, \cdot \rangle$  sobre un espacio vectorial  $\mathbb{V}$  es llamado producto inercial si satisface las siguientes propiedades

1.  $\langle v, v \rangle \geq 0$
2.  $\langle v, v \rangle = 0$  si  $v = 0$
3.  $\langle a_1 v_1 + a_2 v_2, v \rangle = a_1 \langle v_1, v \rangle + a_2 \langle v_2, v \rangle$
4.  $\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle$

para todo  $v, v_1, v_2 \in \mathbb{V}$  y  $a_1, a_2 \in \mathbb{R}$

*Demostración.* La demostración véase [33, p. 31].

**Definición B.0.17.** Un espacio dotado de producto inercial completo es un espacio de Hilbert.

**Definición B.0.18.** Sean  $x_1, x_2$  elementos de un espacio dotado de producto inercial  $\mathbb{X}$  se dice que son ortogonales si  $\langle x_1, x_2 \rangle = 0$ . Una colección contable de elementos  $\{e_1, e_2, \dots\}$  es una sucesión ortonormal si  $\|e_j\| = 1$  para todo  $j$  y además cada par de  $e_j$ 's son ortogonales.

**Teorema B.0.3.** Sean  $\{e_1, e_2, \dots\}$  una sucesión ortonormal en un espacio dotado de producto inercial  $\mathbb{X}$ . Para todo  $x \in \mathbb{X}$ ,  $\sum_{i=1}^{\infty} \langle x, e_i \rangle^2 \leq \|x\|^2$ , y además  $\sum_{i=1}^{\infty} \langle x, e_i \rangle e_i$  converge en  $\mathbb{X}$ .

*Demostración.* La demostración [33, p. 33].

**Definición B.0.19.** Sea  $\{e_n\}$  una sucesión ortonormal en un espacio de Hilbert  $\mathbb{H}$  es una base ortonormal o sistema ortonormal completo (CONS) si  $\overline{\text{span}\{e_n\}} = \mathbb{H}$ .

**Teorema B.0.4.** Una sucesión ortonormal  $\{e_n\}$  en un espacio de Hilbert  $\mathbb{H}$  es un CONS si  $\langle x, e_n \rangle = 0$  para todo  $n$  implica que  $x = 0$ .

*Demostración.* La demostración véase [33, p. 35].

**Teorema B.0.5.** Todo elemento  $x$  de un espacio de Hilbert  $\mathbb{H}$  con CONS  $\{e_j\}$  puede ser expresado en términos de la expansión de Fourier

$$x = \sum_{j=1}^{\infty} \langle x, e_j \rangle e_j \quad (\text{B.1})$$

$$\|x\|^2 = \sum_{j=1}^{\infty} \langle x, e_j \rangle^2 \quad (\text{B.2})$$

**Teorema B.0.6.** Un espacio de Hilbert es separable si tiene una base ortonormal.

*Demostración.* La demostración véase en [33, p. 35].

### B.0.3 Operadores

En esta sección se expondrá brevemente las propiedades de las transformaciones lineales sobre espacios lineales normados.

Sea  $\mathcal{T} : \mathbb{X}_1 \rightarrow \mathbb{X}_2$  una transformación lineal, los espacios asociados a dicha transformación son los siguientes.

$$Dom(\mathcal{T}) = \text{un subconjunto de } \mathbb{X}_1 \text{ donde está definido } \mathcal{T}$$

$$Im(\mathcal{T}) = \{\mathcal{T}x : x \in Dom(\mathcal{T})\}$$

$$Ker(\mathcal{T}) = \{x \in Dom(\mathcal{T}) : \mathcal{T}x = 0\}$$

Asumiremos que  $Dom(\mathcal{T})$  es un espacio lineal, lo que implica que  $Im(\mathcal{T})$ . El rango del operador  $\mathcal{T}$  esta definido por

$$rank(\mathcal{T}) = dim(Im(\mathcal{T}))$$

Para los fines pertinentes, será también interesante no solo estudiar las transformaciones lineales sino también las transformaciones acotadas como se define a continuación.

**Definición B.0.20.** *Supongamos que  $\mathbb{X}_1, \mathbb{X}_2$  son espacios vectoriales normados con normas  $\|\cdot\|_i$  con  $i = 1, 2$ . Una transformación lineal  $\mathcal{T}$  de  $\mathbb{X}_1$  a  $\mathbb{X}_2$  es acotada si existe un constante  $C > 0$  tal que*



$$\|\mathcal{T}x\|_2 \leq C\|x\|_1$$

para todo  $x \in \mathbb{X}_1$ .

**Teorema B.0.7.** *Una transformación lineal entre dos espacios normados es uniformemente continua si es acotada.*

*Demostración.* La demostración véase en la página 63.

Usaremos  $\mathfrak{B}(\mathbb{X}_1, \mathbb{X}_2)$  para denotar el conjunto de todas las transformaciones lineales acotadas (y, por tanto, continuas). Para un espacio normado se define la siguiente norma.

$$\|\mathcal{T}\|_1 = \sup_{x \in \mathbb{X}_1, \|x\|_1=1} \|\mathcal{T}x\|_2 \tag{B.3}$$

Los elementos de  $\mathfrak{B}(\mathbb{X}_1, \mathbb{X}_2)$  son llamados operadores lineales acotados, operadores lineales u operadores. Si  $\mathbb{X}_1 = \mathbb{X}_2 = \mathbb{X}$  la notación para los operadores acotados sobre  $\mathbb{X}$  será  $\mathfrak{B}(\mathbb{X})$ .

**Teorema B.0.8.** *Sean  $\mathbb{X}_1$  y  $\mathbb{X}_2$  dos espacios normados lineales, si  $\mathbb{X}_2$  es completo, entonces  $\mathfrak{B}(\mathbb{X}_1, \mathbb{X}_2)$  con la norma B.3 es un espacio de Banach.*

*Demostración.* La demostración véase [33, p. 109].

**Teorema B.0.9.** *Sean  $\mathbb{X}_1$  y  $\mathbb{X}_2$  espacios de Banach,  $f$  una función Bochner integrable de  $E$  en  $\mathbb{X}_1$  y  $\mathcal{T} \in \mathfrak{B}(\mathbb{X}_1, \mathbb{X}_2)$ . Entonces  $\mathcal{T}f$  es Bochner integrable.*

$$\mathcal{T} \left( \int_E f d\mu \right) = \int_E \mathcal{T}f d\mu$$

*Demostración.* La demostración véase en [33, p. 65].

## B.0.4 Operadores Adjuntos

**Teorema B.0.10.** Sean  $\mathbb{H}_1, \mathbb{H}_2$  espacios de Hilbert dotados de producto inercia  $\langle \cdot, \cdot \rangle_i$ ,  $i=1,2$ . Para todo  $\mathcal{T} \in \mathfrak{B}(\mathbb{H}_1, \mathbb{H}_2)$  existe un único elemento  $\mathcal{T}^*$  de  $\mathfrak{B}(\mathbb{H}_1, \mathbb{H}_2)$  tal que

$$\langle \mathcal{T} x_1, x_2 \rangle_2 = \langle x_1, \mathcal{T}^* x_2 \rangle_1 \quad (\text{B.4})$$

para todo  $x_1 \in \mathbb{H}_1$   $x_2 \in \mathbb{H}_2$ .

*Demostración.* La demostración véase en [33, p. 72].

**Definición B.0.21.** Sea  $\mathcal{T} \in \mathfrak{B}(\mathbb{H}_1, \mathbb{H}_2)$  el único operado  $\mathcal{T}^*$  de  $\mathfrak{B}(\mathbb{H}_1, \mathbb{H}_2)$  tal que satisface B.4 es el adjunto de  $\mathcal{T}$ . Si  $\mathbb{H}_1 = \mathbb{H}_2$ ,  $\mathcal{T}$  es llamado autoadjunto cuando  $\mathcal{T}^* = \mathcal{T}$ .

## B.0.5 Operadores Compactos

**Definición B.0.22.** Una transformación lineal  $\mathcal{T} : \mathbb{X}_1 \rightarrow \mathbb{X}_2$  es compacta si para cualquier sucesión acotada  $\{x_n\} \in \mathbb{X}_1$ ,  $\{\mathcal{T}x_n\}$  tiene una subsucesión convergente en  $\mathbb{X}_2$ .

**Teorema B.0.11.** Los siguientes resultados se aplican entre operadores compactos sobre espacios normados lineales

1. La clausura del rango de cualquier operador compacto es separable.
2. Los operadores de rango finito son compactos.
3. La composición de dos operadores compactos es compacto si cualquiera de los operadores es compacto.
4. El conjunto de operadores compactos sobre un espacio de Banach es cerrado.

*Demostración.* La demostración véase en [33, p. 93].

**Teorema B.0.12.** Sean  $\mathbb{H}_1, \mathbb{H}_2$  espacios de Hilbert y supongamos que  $\mathcal{T} \in \mathfrak{B}(H_1, H_2)$ . Entonces,

1.  $\mathcal{T}$  es compacto si existe una sucesión  $\{\mathcal{T}_n\}$  de operadores finito dimensionales tal que  $\|\mathcal{T}_n - \mathcal{T}\| \rightarrow 0$  cuando  $n \rightarrow \infty$ .
2.  $\mathcal{T}$  es compacto si  $\mathcal{T}^*$  es compacto.

## B.0.6 Operadores de Hilbert Schmidt

**Teorema B.0.13.** Sean  $\mathcal{T}_1, \mathcal{T}_2$  operadores en  $\mathfrak{B}(\mathbb{H}_1, \mathbb{H}_2)$ . Suponga que existe CONSs  $\{e_{1i}\}$ , y  $\{e_{2j}\}$  para  $\mathbb{H}_1$  y  $\mathbb{H}_2$ , respectivamente, tal que

$$\sum_{i=1}^{\infty} (\|\mathcal{T}_1 e_{1i}\|_2^2 + \|\mathcal{T}_2 e_{1i}\|_2^2) < \infty$$

o

$$\sum_{j=1}^{\infty} (\|\mathcal{T}_1^* e_{2j}\|_2^2 + \|\mathcal{T}_2^* e_{2j}\|_2^2) < \infty$$

**Teorema B.0.14.** Un operador de Hilbert Schimidt es compacto

*Demostración.* Véase la demostración en [33, p. 109].

Claramente,  $\mathfrak{B}_{HS}(\mathbb{H}_1, \mathbb{H}_2)$  es un espacio lineal, por ende lo podemos dotar de un producto inercial que se define a continuación

**Definición B.0.23.** El producto inercial de  $\mathcal{T}_1, \mathcal{T}_2 \in \mathfrak{B}_{HS}(\mathbb{H}_1, \mathbb{H}_2)$  es

$$\langle \mathcal{T}_1, \mathcal{T}_2 \rangle_{HS} = \sum_{j=1}^{\infty} \langle \mathcal{T}_1 e_j, \mathcal{T}_2 e_j \rangle_2$$

donde  $\{e_i\}$  es cualquier CONS para  $\mathbb{H}_1$ . La norma HS de  $\mathcal{T}$ ,  $\|\mathcal{T}_{HS}\|$  es la norma dada por el producto inercial HS

$$\|\mathcal{T}\|_{HS} = \left\{ \sum_{j=1}^{\infty} \|\mathcal{T}e_j\|_2^2 \right\}^{\frac{1}{2}}$$

**Teorema B.0.15.** *El espacio lineal  $\mathfrak{B}_{HS}(\mathbb{H}_1, \mathbb{H}_2)$  es un espacio de Hilbert separable cuando está dotado del producto inercial HS. Para cualquier CONS  $\{e_{1i}\}$  y  $\{e_{2j}\}$  para  $\mathbb{H}_1$  y  $\mathbb{H}_2$ , respectivamente,  $\{e_{1i} \otimes_1 e_{2j}\}$  es un CONS de  $\mathbb{H}_1, \mathbb{H}_2$ .*

*Demostración.* La demostración véase en [33, p. 110].

**Definición B.0.24.** *Sea  $\{e_i\}$  un sistema ortonormal completo para  $\mathbb{H}_1$  y  $\mathcal{T} \in \mathfrak{B}(\mathbb{H}_1, \mathbb{H}_2)$ . Si  $\mathcal{T}$  satisface*

$$\sum_{i=1}^{\infty} \|\mathcal{T}e_i\|_2^2 < \infty$$

*Entonces  $\mathcal{T}$  es un operador Hilbert-Schmidt (HS). La colección de los operadores HS  $\mathfrak{B}(\mathbb{H}_1, \mathbb{H}_2)$  es denotado por  $\mathfrak{B}_{HS}(\mathbb{H}_1, \mathbb{H}_2)$ .*

# Bibliografía

- [1] Frédéric Ferraty y Philippe Vieu. *Nonparametric functional data analysis: theory and practice*. Springer Science & Business Media, 2006.
- [2] Walter Andrew Shewhart. *Economic control of quality of manufactured product*. Macmillan y Co Ltd, London, 1931.
- [3] Harold Hotelling. “Multivariate quality control. Techniques of statistical analysis”. En: *McGraw-Hill, New York* (1947).
- [4] Peter Hall y Nader Tajvidi. “Permutation tests for equality of distributions in high-dimensional settings”. En: *Biometrika* 89.2 (2002), págs. 359-374.
- [5] A Mostajeran, N Iranpanah y R Noorossana. “A new bootstrap based algorithm for HotellingT2 multivariate control chart”. En: *Journal of Sciences, Islamic Republic of Iran* 27.3 (2016), págs. 269-278.
- [6] Colosimo, Bianca M y Pacella Massimo. “A comparison study of control charts for statistical monitoring of functional data”. En: *International Journal of Production Research* 48.6 (2010), págs. 1575-1601.
- [7] Shey-Huei Sheu, Chih-Wen Ouyoung y Tsung-Shin Hsu. “Phase II statistical process control for functional data”. En: *Journal of Statistical Computation and Simulation* 83.11 (2013), págs. 2144-2159.
- [8] D.R. García. “Cartas de control para datos funcionales”. Tesis de mtría. CIMAT, 2011.
- [9] O.P.Rodrigo. “Monitoreo de datos Funcionales”. Tesis de mtría. CIMAT, 2013.
- [10] Douglas C Montgomery. *Introduction to statistical quality control*. John Wiley & Sons, 2020.

- [11] Douglas C Montgomery y Dirk Valckx Verbeeck. *Control estadístico de la calidad*. 658.562 M66 2004. Limusa Wiley, 2004.
- [12] Sergio Yáñez Canal, José Alberto Vargas Navas y González Álvarez Nelfi Gertrudis. “Cartas de control T2 multivariadas usando R y SAS”. En: ().
- [13] Robert L Mason y John C Young. *Multivariate statistical process control with industrial applications*. SIAM, 2002.
- [14] P.J. Green y B.W. Silverman. *Nonparametric regression and generalized linear models: a roughness penalty approach*. English. United Kingdom: Chapman y Hall, 1994. ISBN: 0412300400.
- [15] Ronald B Crosier. “Multivariate generalizations of cumulative sum quality-control schemes”. En: *Technometrics* 30.3 (1988), págs. 291-303.
- [16] Cynthia A Lowry, William H Woodall, Charles W Champ y Steven E Rigdon. “A multivariate exponentially weighted moving average control chart”. En: *Technometrics* 34.1 (1992), págs. 46-53.
- [17] Willis A Jensen, L Allison Jones-Farmer, Charles W Champ y William H Woodall. “Effects of parameter estimation on control chart properties: a literature review”. En: *Journal of Quality technology* 38.4 (2006), págs. 349-364.
- [18] Eugenio K Epprecht. “Statistical control of multiple-stream processes: A literature review”. En: *Frontiers in Statistical Quality Control 11* (2015), págs. 49-64.
- [19] Søren Bisgaard. “The future of quality technology: From a manufacturing to a knowledge economy & from defects to innovations”. En: *Quality Engineering* 24.1 (2012), págs. 30-36.
- [20] Yajun Mei. “Efficient scalable schemes for monitoring a large number of data streams”. En: *Biometrika* 97.2 (2010), págs. 419-433.
- [21] William H Woodall y Douglas C Montgomery. “Some current directions in the theory and application of statistical process monitoring”. En: *Journal of Quality Technology* 46.1 (2014), págs. 78-94.

- 
- [22] Lee J Wells, Fadel M Megahed, Jaime A Camelio y William H Woodall. “A framework for variation visualization and understanding in complex manufacturing systems”. En: *Journal of Intelligent Manufacturing* 23.5 (2012), págs. 2025-2036.
- [23] Fadel M Megahed y L Allison Jones-Farmer. “Statistical perspectives on “big data””. En: *Frontiers in statistical quality control* 11. Springer, 2015, págs. 29-47.
- [24] James O Ramsay y Bernard W Silverman. *Applied functional data analysis: methods and case studies*. Vol. 77. Springer, 2002.
- [25] Dan J Spitzner, JS Marron y GK Essick. “Mixed-model functional ANOVA for studying human tactile perception”. En: *Journal of the American Statistical Association* 98.462 (2003), págs. 263-272.
- [26] JA Cuesta-Albertos y Manuel Febrero-Bande. “A simple multiway ANOVA for functional data”. En: *Test* 19.3 (2010), págs. 537-557.
- [27] Alessia Pini, Aymeric Stamm y Simone Vantini. “Hotelling’s T2 in separable Hilbert spaces”. En: *Journal of Multivariate Analysis* 167 (2018), págs. 284-305.
- [28] R Cao Abad R & Fenandez Casal. “Tecnicas de Remuestreo”. En: (2017).
- [29] RA Fisher. “The Design of Experiments (Hafner, New York)”. En: (1935).
- [30] Edwin JG Pitman. “Significance tests which may be applied to samples from any populations”. En: *Supplement to the Journal of the Royal Statistical Society* 4.1 (1937), págs. 119-130.
- [31] Edwin James George Pitman. “Significance tests which may be applied to samples from any populations: III. The analysis of variance test”. En: *Biometrika* 29.3/4 (1938), págs. 322-335.
- [32] Manuel Febrero, Pedro Galeano y Wenceslao González-Manteiga. “Outlier detection in functional data by depth measures, with application to identify abnormal NOx levels”. En: *Environmetrics: The official journal of the International Environmetrics Society* 19.4 (2008), págs. 331-345.
- [33] Salomon Bochner. “Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind”. En: *Fundamenta Mathematicae* 20.1 (1933), págs. 262-176.

- [34] Tailen Hsing y Randall Eubank. *Theoretical foundations of functional data analysis, with an introduction to linear operators*. Vol. 997. John Wiley & Sons, 2015.
- [35] Piercesare Secchi, Aymeric Stamm y Simone Vantini. “Inference for the mean of large  $p$  small  $n$  data: A finite-sample high-dimensional generalization of Hotelling’s theorem”. En: *Electronic journal of statistics* 7 (2013), págs. 2005-2031.
- [36] James O Ramsay y CJ Dalzell. “Some tools for functional data analysis”. En: *Journal of the Royal Statistical Society: Series B (Methodological)* 53.3 (1991), págs. 539-561.
- [37] Lajos Horváth y Piotr Kokoszka. *Inference for functional data with applications*. Vol. 200. Springer Science & Business Media, 2012.
- [38] B. Efron. “Bootstrap Methods Another Look at the Jackknife”. En: *The Annals of Statistics* 7.1 (1979), págs. 1-26.
- [39] Eric W Noreen. *Computer intensive methods for testing hypotheses*. John Wiley & Sons, 1989.
- [40] Emanuel Parzen. “On estimation of a probability density function and mode”. En: *The annals of mathematical statistics* 33.3 (1962), págs. 1065-1076.
- [41] M Rosenblatt. “Remarks on some nonparametric estimates of a density function. Annals of Mathematical Statistics”. En: (1956).
- [42] Rosa Baranzano. *Non-parametric kernel density estimation-based permutation test: Implementation and comparisons*. 2011.
- [43] G Fraiman R. y Muniz. “Trimmed means for functional data”. En: *Sociedad de Estadística e Investigación Operativa* 10.2 (2001).
- [44] Antonio Cuevas, Manuel Febrero y Ricardo Fraiman. “On the use of the bootstrap for estimating functions with functional data”. En: *Computational statistics & data analysis* 51.2 (2006), págs. 1063-1074.
- [45] Antonio Cuevas, Manuel Febrero y Ricardo Fraiman. “On the use of the bootstrap for estimating functions with functional data”. En: *Computational statistics & data analysis* 51.2 (2006), págs. 1063-1074.



- 
- [46] Antonio Cuevas, Manuel Febrero y Ricardo Fraiman. “Robust estimation and classification for functional data via projection-based depth notions”. En: *Computational Statistics* 22.3 (2007), págs. 481-496.
- [47] Juan Antonio Cuesta-Albertos y Alicia Nieto-Reyes. “The random Tukey depth”. En: *Computational Statistics & Data Analysis* 52.11 (2008), págs. 4979-4988.
- [48] WHO / Metrics: Disability-Adjusted Life Year (DALY). URL: <https://github.com/astamm/fdahotelling/tree/master/data>. (visitado 17-12-2019).
- [49] Laura M Sangalli, Piercesare Secchi, Simone Vantini y Alessandro Veneziani. “A case study in exploratory functional data analysis: geometrical features of the internal carotid artery”. En: *Journal of the American Statistical Association* 104.485 (2009), págs. 37-48.
- [50] Tiziano Passerini, Laura M Sangalli, Simone Vantini, Marina Piccinelli, Susanna Bacigaluppi, Luca Antiga, Edoardo Boccardi, Piercesare Secchi y Alessandro Veneziani. “An integrated statistical investigation of internal carotid arteries of patients affected by cerebral aneurysms”. En: *Cardiovascular Engineering and Technology* 3.1 (2012), págs. 26-40.
- [51] Laura M Sangalli, Piercesare Secchi y Simone Vantini. “AneuRisk65: A dataset of three-dimensional cerebral vascular geometries”. En: *Electronic Journal of Statistics* 8.2 (2014), págs. 1879-1890.
- [52] Jinyuan Chang, Wen Zhou, Wen-Xin Zhou y Lan Wang. “Comparing large covariance matrices under weak conditions on the dependence structure and its application to gene clustering”. En: *Biometrics* 73.1 (2017), págs. 31-41.
- [53] Hermine Biermé. “Introduction to random fields and scale invariance”. En: *Stochastic geometry*. Springer, 2019, págs. 129-180.