

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

DISEÑO DE GRÁFICOS DE CONTROL NO PARAMÉTRICOS PARA DATOS AUTOCORRELACIONADOS E INDEPENDIENTES UTILIZANDO TÉCNICAS DE REMUESTREO

TRABAJO PREVIO A LA OBTENCIÓN DEL TÍTULO DE INGENIERO
MATEMÁTICO

PROYECTO DE INVESTIGACIÓN

DANIEL MATEO LARCO ALVAREZ
mateo_dan@hotmail.com

Director: MIGUEL ALFONSO FLORES SANCHEZ.
miguel.flores@epn.edu.ec

QUITO, NOVIEMBRE 2018

DECLARACIÓN

Yo DANIEL MATEO LARCO ALVAREZ, declaro bajo juramento que el trabajo aquí escrito 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.



Daniel Mateo Larco Alvarez

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por DANIEL MATEO LARCO ALVAREZ, bajo mi supervisión.

A handwritten signature in blue ink, appearing to read 'Miguel Alfonso Flores Sanchez', written over a horizontal line.

Miguel Alfonso Flores Sanchez.
Director del Proyecto

AGRADECIMIENTOS

A Dios por haberme guiado toda mi vida y haberme permitido elegir esta hermosa carrera.

A mis padres por su apoyo incondicional en cada cosa que he realizado.

A Joss por alegrar mi días, ser participe de mis sueños y anhelos y estar a mi lado en los momentos que más le necesité.

A mis amigos, Katherine y Paúl con quienes compartí durante muchos años no solo estudio sino amistad y compañerismo.

A Miguel, por su guía en el desarrollo del trabajo de titulación y su confianza en mis capacidades.

DEDICATORIA

A mis padres y Joss por su confianza, apoyo y amor incondicional.

Índice general

Índice de figuras	VIII
Índice de cuadros	X
Resumen	XIV
Abstract	XVI
Prólogo	XVII
Objetivos	XXI
1. Introducción	1
1.1. Antecedentes	1
1.2. Control estadístico de la calidad y metodología Seis Sigma	5
1.2.1. Calidad	5
1.2.2. Control estadístico de procesos	5
1.2.3. Metodología Seis Sigma	6
1.3. Gráficos de control	10
1.3.1. Gráficos de control y pruebas de hipótesis	11
1.3.2. Fase I y Fase II de un gráfico de control	14
1.3.3. Tipos de gráficos de control	16
2. Marco teórico	25
2.1. Profundidad de datos	25
2.1.1. Introducción	25

2.1.2.	Conceptos básicos	26
2.1.3.	Medidas de profundidad	28
2.2.	Técnicas de remuestreo	31
2.2.1.	Método Bootstrap	32
2.2.2.	Bootstrap Suavizado	33
2.2.3.	Bootstrap estacionario	35
2.3.	Gráficos de control basados en la profundidad de datos	38
2.3.1.	Estadísticos derivados de la profundidad de datos	39
2.3.2.	Gráficos de control	41
2.3.3.	Simulación	43
2.4.	Gráficos de control basados en métodos de remuestreo	46
2.4.1.	Gráfico para muestras i.i.d. basadas en el método bootstrap	46
2.4.2.	Gráficos de control para observaciones correlacionadas	48
2.4.3.	Gráficos de control usando SB	51
3.	Propuesta de gráficos de control para la Fase I y II	54
3.1.	Metodología	55
3.1.1.	Gráficos de control para calibrado	55
3.1.2.	Gráficos de control para monitoreo	58
3.2.	Estudio de simulación	60
3.2.1.	Evaluación de desempeño para gráfico de control de calibrado	61
3.2.2.	Evaluación de desempeño gráfico de monitoreo	76
3.3.	Caso de estudio de extrusión de palitos de chupete	95
4.	Conclusiones y Recomendaciones	108
	Bibliografía	111
	Anexos	115
A.1.	Código en R para simulaciones propuestas por Liu	115
A.1.1.	Simulación de gráficos de control para procesos multivariantes	115
A.1.2.	Gráficos de control basados en métodos bootstrap	118

A.2. Código en R para simulación observaciones independientes y dependientes	124
A.2.1. Funciones para cálculo de estadísticos	124
A.2.2. Simulación Fase I caso independiente	147
A.2.3. Simulación Fase I caso dependiente	149
A.2.4. Simulación Fase II caso independiente	152
A.2.5. Simulación Fase II caso dependiente	155
A.2.6. Tablas para Fase II, variación de un componente del vector de medias	159
A.2.7. Tablas para Fase II, variación todos los componentes del vector de medias	160
A.2.8. Gráficos de control para Fases I y II	162
A.2.9. Aplicación a caso de estudio extrusión de palitos, caso independiente	168
A.2.10. Aplicación caso de estudio extrusión de palitos, caso dependiente	171

Índice de figuras

1.1. Proceso mejora proceso	6
1.2. Variable normal centrada en $\mu = T$, porcentaje de elementos no conformes segun el criterio Seis Sigma (capacidad de procesos). Se muestran los defectos por millon (ppm Defective) esperados para cada nivel de calidad del producto o capacidad de proceso.	7
1.3. Etapas y técnicas asociadas al proceso DMAIC	8
1.4. Proceso de monitoreo de una variable que afecta la calidad de un producto mediante la aplicación de gráficos de control	15
1.5. Clasificación de gráficos de control de acuerdo a la variable CTQ	16
1.6. Gráfico de control \bar{x} usando el rango como medida de variabilidad	18
1.7. Gráfico de control R	19
1.8. Gráfico de control \bar{x} usando la desviación estándar como medida de variabilidad	20
1.9. Gráfico de control S	21
1.10. Gráfico de control $CUSUM$	22
1.11. Gráfico de control $EWMA$	23
1.12. Clasificación de gráficos de control de acuerdo al tipo de muestra y distribución de las observaciones	24
2.1. Ilustración procedimiento de convex hull peeling	27
2.2. Gráfico r	44
2.3. Gráfico Q para $n=4$	45
2.4. Gráfico Q para $n=10$	45
2.5. Comparación bajo distribución normal ($\alpha = 5\%$)	47
2.6. Comparación bajo distribución exponencial ($\alpha = 5\%$)	48

2.7. Gráfico X-barra para datos dependientes	51
3.1. Cambio en la media del proceso	62
3.2. Palito de chupete	95
3.3. Proceso mezclado de insumos	96
3.4. Componentes extrusora	96
3.5. Componentes extrusora II	97
3.6. Componentes extrusora II	97
3.7. Características de calidad	98
3.8. Gráfico T^2 de Hotelling para la Fase I, observaciones independientes	100
3.9. Gráfico de control para Fase I observaciones independientes	101
3.10. Gráfico T^2 de Hotelling para la Fase II, observaciones independientes	102
3.11. Gráfico de profundidad:muestra de referencia vs monitoreo para el caso independiente	103
3.12. Gráfico T^2 de Hotelling para la Fase I, observaciones correlacionadas .	104
3.13. Gráfico de control para Fase I observaciones dependientes	105
3.14. Gráfico T^2 de Hotelling para la Fase II, observaciones dependientes .	106
3.15. Gráfico de profundidad:muestra de referencia vs monitoreo para el caso dependiente	107

Índice de cuadros

1.1. Tipos de errores estadísticos	12
3.1. Fase I, Caso independiente para $n = 15$	64
3.2. Fase I, Caso independiente para $n = 25$	65
3.3. Fase I, Caso independiente para $n = 50$	65
3.4. Fase I, Caso independiente para $n = 100$	65
3.5. Estimación de potencia de la prueba para $n = 15$ para caso independiente	66
3.6. Estimación de potencia de la prueba para $n = 25$ para caso independiente	67
3.7. Estimación de potencia de la prueba para $n = 50$ para caso independiente	67
3.8. Estimación de potencia de la prueba para $n = 100$ para caso independiente	68
3.9. Resultados para el caso dependiente de la estimación de la tasa de falsas alarmas para $n = 15$	69
3.10. Resultados para el caso dependiente de la estimación de la tasa de falsas alarmas para $n = 25$	69
3.11. Resultados para el caso dependiente de la estimación de la tasa de falsas alarmas para $n = 50$	70
3.12. Resultados para el caso dependiente de la estimación de la tasa de falsas alarmas para $n = 100$	70
3.13. Estimación de potencia de la prueba para $n = 15$ para caso dependencia	72
3.14. Estimación de potencia de la prueba para $n = 25$ para caso dependencia	73
3.15. Estimación de potencia de la prueba para $n = 50$ para caso dependencia	74

3.16. Estimación de potencia de la prueba para $n = 100$ para caso de dependencia	75
3.17. Potencia del gráfico de control $1 - \beta$ variando un componente del vector de medias para $n = 15$ para escenario de independencia . . .	78
3.18. Potencia del gráfico de control $1 - \beta$ variando un componente del vector de medias para $n = 25$ para escenario de independencia . . .	79
3.19. Potencia del gráfico de control $1 - \beta$ variando un componente del vector de medias para $n = 50$ para escenario de independencia . . .	80
3.20. Potencia del gráfico de control $1 - \beta$ variando un componente del vector de medias para $n = 100$ para escenario de independencia . . .	81
3.21. Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 15$ para escenario de independencia .	82
3.22. Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 25$ para escenario de independencia .	83
3.23. Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 50$ para escenario de independencia .	84
3.24. Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 100$ para escenario de independencia .	85
3.25. Potencia del gráfico de control $1 - \beta$ variando una componente del vector de medias para $n = 15$ para escenario de dependencia	87
3.26. Potencia del gráfico de control $1 - \beta$ variando una componente del vector de medias para $n = 25$ para escenario de dependencia	88
3.27. Potencia del gráfico de control $1 - \beta$ variando una componente del vector de medias para $n = 50$ para escenario de dependencia	89
3.28. Potencia del gráfico de control $1 - \beta$ variando una componente del vector de medias para $n = 100$ para escenario de dependencia	90
3.29. Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 15$ para escenario de dependencia . .	91
3.30. Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 25$ para escenario de dependencia . .	92
3.31. Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 50$ para escenario de dependencia . .	93

3.32. Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 100$ para escenario de dependencia . . . 94

Resumen

El gráfico de control es una de las herramientas principales en el control estadístico de procesos que permite recopilar y analizar datos con la finalidad de evaluar, monitorear y controlar el rendimiento de procesos de manufactura. Mediante el uso de los gráficos de control se pretende identificar variaciones en el proceso debido a causas comunes y causas especiales.

La propuesta del diseño de un gráfico de control para observaciones independientes y dependientes que permitan controlar la variabilidad de procesos y servicios se basa en la detección de anomalías mediante técnicas de remuestreo bootstrap y adaptación de gráficos no paramétricos de profundidades para medidas de calidad multivariantes y univariantes, se propone la aplicación de métodos con el objeto de eliminar atípicos basados en la profundidad de datos para la eliminación de las principales causas asignables de la variabilidad del proceso, causante de errores y anomalías también denominada Fase I. A continuación, a partir de los valores de profundidad de datos de la muestra de calibrado resultante, se propone la aplicación del gráfico no paramétrico de profundidades para llevar a cabo la Fase II o de monitoreo, es decir, el contraste de si una nueva observación pertenece a la población estimada en la Fase I. Esta propuesta presenta ciertas ventajas sobre los gráficos tradicionales debido a que detectan simultáneamente el cambio de localización e incremento de escala del proceso.

Los gráficos de control propuestos se comparan con los gráficos planteados por Regina Liu [Liu, 1995] para la cual se obtienen mejores resultados en base a potencia para los distintos tamaños de muestra simulados, además los gráficos de control propuestos son consistentes, es decir, cuando aumenta el tamaño de la muestra también aumenta la potencia de la prueba.

Finalmente, se aplica la metodología desarrollada para el caso de estudio de extrusión de palitos fabricados en Confiteca C.A. de donde se obtiene una mejor capacidad de detección de atípicos para las Fases I y II en comparación a gráficos tradicionales.

Palabras clave: control estadístico de procesos, gráficos de control, no paramétrico, profundidad de datos, bootstrap, bootstrap estacionario, ARL.

Abstract

The control chart is one of the principal tools in statistical process control that allows to collect and analyze data in order to evaluate, monitor and control the performance of manufacture processes. The use of the control chart aims to identify process shifts due to common and special causes.

The proposal of the design of a control chart for independent and dependent observations that allows controlling the variability of processes and services is based on the detection of anomalies by means of bootstrap re sampling techniques and adaptation of non-parametric depth plots for multivariate and univariate quality measurements, the application of methods is proposed in order to eliminate atypical data based on data depth for the elimination of the main assignable causes of the variability of the process, causing errors and abnormalities also called Phase I. Then, from the data depth values of the resulting calibration sample, the application of the nonparametric depth chart is proposed to carry out Phase II or monitoring, that is, the contrast of whether a new observation belongs to the population estimated in the Phase I. This proposal has certain advantages over traditional graphics because they simultaneously detect the change of location and scale increase of the process. The proposed control charts are compared with the charts proposed by Regina Liu [Liu, 1995] for which better results are obtained based on the power test for the different simulated sample sizes, in addition the proposed control charts are consistent, that is, When the sample size increases, the power of the test also increases.

Finally, the methodology developed for the case study of extrusion of sticks manufactured in Confiteca C.A. from which a better atypical detection capacity is obtained for Phases I and II compared to traditional graphs.

Keywords: statistical process control, control charts, nonparametric, data depth, bootstrap, stationary bootstrap, ARL.

Prólogo

A lo largo de las últimas décadas ha aumentado el interés por medir, mejorar y controlar la calidad de los productos, servicios y procedimientos. Esto está conectado a la fuerte relación entre calidad, productividad, prestigio, confianza e imagen de marca. Estos procedimientos de implementación del control estadístico de la calidad (CEC) están actualmente relacionados con el aumento de la competitividad de las empresas.

Se puede definir al control estadístico de la calidad como una técnica de analizar procesos, fijar estándares, comparar rendimientos, verificar y estudiar desviaciones, para buscar e implementar soluciones. La aplicación de esta técnica a un proceso consta de cinco etapas: Definición, Medida, Análisis, Mejora y Control, las siglas DMAIC provienen de las correspondientes palabras inglesas [Breyfogle, 1999]. Se podría decir que el ciclo DMAIC se basa en Definir los problemas y situaciones a mejorar, Medir para obtener la información, Analizar la información recogida, Incorporar mejoras en los procesos y Controlar los procesos o productos existentes. La etapa de control se caracteriza por el uso de herramientas basadas en la detección y corrección de errores, el gráfico de control es una de las herramientas principales en el control estadístico de procesos que permite recopilar y analizar datos con la finalidad de evaluar, monitorear y controlar el rendimiento de procesos de manufactura. Mediante el uso de los gráficos de control se pretende identificar variaciones en el proceso debido a causas comunes y causas especiales, ya que ningún proceso se encuentra de forma espontánea bajo control, conseguir y mantener esto supone una mejora sustancial al proceso.

Para controlar la variabilidad general de un producto, puede ser necesario tener gráficos de control que puedan monitorear medidas de calidad multivariantes directamente. La calidad de un producto la definen varias variables de forma simultánea, de manera que si alguna de ellas no cumple con los requisitos de calidad el producto es no conforme en su totalidad. Por este punto, puede ser esencial identificar un estándar común o singular de la calidad general de productos y, consecuentemente

determinar la variabilidad aleatoria alrededor de este objetivo.

A pesar de que los gráficos de control son fáciles de construir, interpretar, visualizar y que en general son efectivos en la práctica, están diseñados para medidas de calidad univariantes y por lo general están basados en supuestos paramétricos para modelar el comportamiento del proceso, los cuales constituyen un limitante en su aplicación pues en caso de no verificarse, la validez del método queda afectada al igual que las conclusiones que se derivan del mismo. Muchas industrias aplican procedimientos de control estadístico de procesos, sin verificar previamente los supuestos teóricos, la aplicación de técnicas estándar cuando se viola las hipótesis conlleva a que se genere falsas alarmas que describen de manera errónea el proceso.

Para el caso multivariante, se han desarrollado gráficos de control no paramétricos tales como los propuestos por Liu [1995] en los cuales se asume independencia entre observaciones. Este enfoque no paramétrico permite abordar este problema mediante el concepto de profundidad de datos, la definición de profundidad está estrictamente ligada a los estadísticos de orden. En el caso multivariante permite contar con una noción de orden del centro hacia afuera.

En este caso se busca dar una idea de centralidad en los datos, la profundidad de un punto relativa a un conjunto de datos mide que tan profundo se halla ese punto dentro de la nube de datos y proporciona un orden de puntos del centro hacia afuera en cualquier dimensión.

A partir de esta medida de profundidad la propuesta de Liu se basa en la distribución asintótica del estadístico Rank que no depende de la distribución original de la característica de calidad. Las observaciones originales son reducidas a rangos y las nuevas observaciones son controladas comparando su profundidad respecto a la profundidad de las observaciones supuestamente en control. Sin embargo, cuando el tamaño de muestra es pequeña, por el acercamiento asintótico del estadístico, el gráfico de control tiene poca potencia.

Otra de las hipótesis de partida para la implementación de gráficos de control tipo Shewhart asume que las observaciones no presentan una estructura de autocorrelación. Sin embargo, con el aumento de la automatización, la información acerca del estado de los procesos se obtiene a tasas mayores, esto significa intervalos más cortos de medición que son más propensos a la presencia de autocorrelación en las observaciones. Esta condición está presente en los datos generados por procesos continuos y de producción por lotes, donde el valor particular del parámetro u observación en el tiempo depende de los valores previos.

La independencia entre observaciones implica que las observaciones actuales no

proporcionan ninguna información del siguiente valor de los datos. Un proceso que viola el requisito de independencia con frecuencia causaría que la variación a corto plazo sea un indicador deficiente de la variación a largo plazo. El principal efecto de utilizar datos autocorrelacionados para realizar control estadístico es que genera límites de control estrechos, lo que provoca un incremento de la tasa de falsas alarmas y una disminución en la capacidad de detección. Una idea sencilla para evitar la autocorrelación consiste en el aumento del intervalo de muestreo; sin embargo, el uso ineficiente de los datos disponibles puede causar una disminución en la potencia de los gráficos.

Para el caso dependiente, Liu and Tang [1996] proponen una alternativa para el diseño de gráficos de control para estadísticos tradicionales tipo Shewhart cuando no se cumple el supuesto paramétrico en las observaciones, la construcción de los límites de control está basado en el método bootstrap por bloques que consiste en la extracción de bloques de tamaño m de observaciones en el tiempo, al retener las observaciones vecinas juntas dentro de los bloques, se preserva la estructura de dependencia de las variables aleatorias a cortas distancias de retardo. Como resultado, los bloques de remuestreo le permiten a uno llevar esta información a las variables bootstrap. Sin embargo, los límites de control son válidos cuando la distribución muestral del estadístico en estudio es la misma a través del tiempo.

Politis and Romano [1994] generalizan este procedimiento realizando remuestreo de "bloques de bloques" de observaciones para obtener procedimientos asintóticos válidos para parámetros multivariantes de distribución conjunta de series de tiempo estacionarias. El Bootstrap estacionario de manera similar al Bootstrap por bloques involucra el remuestreo de los datos originales para formar una pseudo serie temporal del cual el estadístico o cantidad de interés puede ser recalculado. Este procedimiento es repetido para construir una aproximación de la distribución muestral del estadístico. Sin embargo, en contraste con los métodos por bloques mencionados anteriormente, las pseudo series temporales generadas por el Bootstrap estacionario son en realidad una serie temporal estacionaria. Así, este procedimiento trata de replicar el modelo original al retener la propiedad de estacionariedad en las pseudo series remuestreadas.

La propuesta del diseño de un gráfico de control para datos independientes y dependientes se basa en la detección de puntos atípicos mediante el cálculo de profundidad de datos, remuestreo Bootstrap y la adaptación de los gráficos no paramétricos de profundidades. En particular, se propone una metodología de control que proporciona herramientas para llevar a cabo la Fase II de un gráfico de control. Es

decir, se propone la aplicación de métodos con el objeto de eliminar atípicos basados en la profundidad de datos para la eliminación de las principales causas asignables de la variabilidad del proceso, causante de errores y anomalías también denominada Fase I. A continuación, a partir de los valores de profundidad de datos de la muestra de calibrado resultante, se propone la aplicación del gráfico no paramétrico de profundidades para llevar a cabo la Fase II o de monitoreo, es decir, el contraste de si una nueva observación pertenece a la población estimada en la Fase I.

Los capítulos 1 y 2 corresponden a las secciones de introducción y marco teórico respectivamente. El capítulo 1 abarca los conceptos principales relacionados al control de calidad tales como el control estadístico de los procesos y la metodología Seis sigma, además presenta una introducción y bases estadísticas de los gráficos de control tradicionales así como también ejemplos de gráficos de control. Por otro lado el capítulo 2 engloba las herramientas teóricas como fundamento del trabajo de titulación tales como profundidad de datos y técnicas de remuestreo así como gráficos de control que han desarrollado diversos autores mediante el uso de tales técnicas.

En el capítulo 3 del presente trabajo de titulación se plantea el diseño de gráficos de control no paramétricos utilizando técnicas de remuestreo para la Fase II del control estadístico de procesos, que permitan controlar la variabilidad de procesos y servicios, y así detectar anomalías, a partir de datos autocorrelacionados e independientes. El diseño de los gráficos de control se basa en una adaptación de la detección de datos atípicos en análisis de datos funcionales (Flores, 2018), la cual busca construir gráficos de control para datos funcionales basados en la detección de atípicos a partir del cálculo de la profundidad de datos funcionales propuesta por Febrero (2008) y Hyndman (2010), por un lado, y en el gráfico de control de rangos o r propuesto por Regina Liu para datos multivariantes.

En el capítulo 4 se evaluará el rendimiento y desempeño del procedimiento propuesto mediante un estudio de simulación y su aplicación a datos reales para el caso de estudio de extrusión de palitos de chupete fabricados en Confiteca C.A.

Finalmente en el capítulo 5, se desarrollarán las principales conclusiones y recomendaciones.

Objetivos

El objetivo del presente trabajo de investigación se basa en diseñar gráficos de control no paramétricos basado en profundidad de datos y técnicas de remuestreo, que permitan monitorear datos dependientes e independientes multivariantes.

Objetivos Específicos

En base al objetivo general de este trabajo se plantean los siguientes objetivos específicos:

- Diseñar gráficos de control similares a los convencionales (Shewhart) para la Fase II del control estadístico de procesos para datos autocorrelacionados e independientes.
- Realizar un estudio de simulación para evaluar el rendimiento de los gráficos de control para la Fase II, donde se consideran diferentes tamaños de la muestra de calibrado y de monitoreo, así como diferentes profundidades de datos.
- Implementar los métodos mediante el desarrollo de funciones en R para facilitar el uso de los gráficos de control.
- Aplicar los gráficos de control diseñados en un contexto industrial.

Capítulo 1

Introducción

1.1. Antecedentes

El presente capítulo tiene como objetivo presentar una recopilación bibliográfica sobre el estado del arte de los métodos actuales para el tratamiento de datos autocorrelacionados y medidas de calidad multivariantes, esto conllevará al conocimiento respecto a cómo diferentes autores han abordado esta problemática, los resultados que han obtenido y los problemas aún persistentes en su desarrollo.

Además, se aborda los conceptos teóricos que permitirán el diseño de los gráficos de control propuestos, se comenzará con una introducción al concepto de calidad, el cual dará paso al control estadístico de la calidad (CEC) y a la metodología Seis Sigma, y se presentará una introducción a los gráficos de control estándar (construcción, tipos de gráficos de control)

Shewhart (1924) define a los gráficos de control como herramientas que dentro del control estadístico de procesos permiten monitorear características de calidad de un proceso industrial o de servicios y son ampliamente usados en la actualidad. Un proceso bajo control es aquel que está libre de causas especiales o asignables de variación, es decir, un proceso que se comporta igualmente en el tiempo. Pese a que, gráficos de control convencionales tales como Shewhart [1931], CUSUM (sumas acumuladas) [Page, 1961], y EWMA (promedios móviles ponderados exponencialmente) [Roberts, 1959] son efectivos en la práctica, están basados en las suposiciones que los datos generados por el proceso cuando está en control son independientes y normalmente distribuidos [Alwan and Roberts, 1988].

En procesos de manufactura donde se realiza un monitoreo a través de dos o más características de calidad, no es recomendable el uso de gráficos de control in-

individuales para monitorear variables por separado y luego aplicar la desigualdad de Bonferroni para lograr cotas del nivel de la prueba combinada. En Alt [1982] se muestra que esta desigualdad no es lo suficientemente buena para dar un buen nivel de precisión, a menos que las variables sean independientes. Más precisamente, este acercamiento tiende a sobreestimar la probabilidad de afirmar que el proceso está en control.

Shewhart [1931] consideró para el caso univariante el efecto de no-normalidad en un gráfico de control, específicamente el gráfico \bar{x} . Distribuciones no normales no solo reducen la precisión del gráfico de control tradicional, pero Jones [2002] mostró que la estimación de parámetros supuesta puede afectar en gran manera el rendimiento del gráfico de control.

Del mismo modo, se encuentran métodos para construir gráficos de control multivariantes en la literatura, métodos conocidos incluyen el gráfico T^2 de Hotelling [Hotelling, 1947] que supone una distribución normal multivariante para el conjunto de variables en estudio, y que exige que se conozca tanto el vector de medias como la matriz de varianzas y covarianzas.

Asimismo, se incluyen el gráfico MCUSUM [Woodall and Ncube, 1985] y gráfico MEWMA [Lowry et al., 1992] los cuales utiliza la información contenida en todas las variables consideradas reduciéndolas a un único escalar que las resume y representa, y posteriormente a este escalar le aplica un gráfico CUSUM univariante y EWMA univariante respectivamente.

No obstante, estos gráficos al igual que en el caso univariante parten de un supuesto paramétrico y no consideran autocorrelación entre las variables.

Existen propuestas introducidas por algunos autores ([Scepi, 2002], [Scepi and Acconcia, 1996]), los cuales son derivados usando conjuntamente algoritmos de remuestreo no paramétricos y técnicas de datos factoriales multidimensionales.

Para el caso multivariante, Liu [1995] propone un método no paramétrico basado en la profundidad de datos para graficar mediciones independientes. Estos gráficos pueden ser visualizados tan fácilmente como los gráficos de control univariantes estándar, y pueden detectar simultáneamente el cambio de media y el incremento de escala de un proceso.

Otra de las hipótesis de partida para la implementación de gráficos de control tipo Shewhart parte del hecho que las observaciones son independientes, esta suposición de independencia será incumplida de forma cada vez más frecuente en la práctica, en algunas ocasiones, por la presencia de equipos de medición más sofisti-

cados, que permiten obtener un mayor volumen de información en un intervalo de tiempo más pequeño, favoreciendo la aparición de autocorrelación en las observaciones. En otros casos, la autocorrelación va a ser un comportamiento inherente al proceso, que debe ser asumido como un elemento más en su modelización y análisis. La presencia de autocorrelación tiene un serio impacto sobre el incremento sustancial en la frecuencia de falsas alarmas en dichos gráficos. Estos niveles de autocorrelación entre observaciones sucesivas pueden tener efectos negativos en las propiedades estadísticas de los gráficos de control tradicionales. [Psarakis and Papaleonida, 2007]

Varios autores ([Alwan and Roberts, 1988],[Montgomery and Mastrangelo, 1991], [Apley and Tsung, 2002], [Kim et al., 2010]) han desarrollado métodos alternativos que presentan mejor desempeño que los métodos tradicionales para manejar los efectos de la autocorrelación. Uno de los métodos más comunes es el uso de modelos de series de tiempo, que ajusta los datos a un modelo adecuado y aplica gráficos de control a los residuales. Si el modelo es exacto, entonces los residuos del modelo son independientes. Sin embargo, muchos autores concuerdan que los gráficos de residuos no tienen las mismas propiedades que los gráficos de las observaciones originales. Más aún, si el modelo no es adecuado, los residuos no son independientes idénticamente distribuidos. Jing and Changchun [2002] señaló que el análisis de residuos es insensible a la media cuando el proceso está autocorrelacionado positivamente.

Otro acercamiento utiliza ajuste de límites de control. Wardell et al. [1994] señaló que el hecho de que los límites de control de un gráfico de control estarían afectados por autocorrelación entre datos. Para procesos autocorrelacionados positivamente, los límites de control estarán más ajustados que lo deseado, lo que provoca un incremento sustancial en la tasa promedio de falsas alarmas. Para procesos autocorrelacionados negativamente, los límites de control son mucho más anchos de lo esperado, lo que provoca una disminución en la capacidad de detectar cambios en el proceso. Entonces los límites de control deberían ser ajustados para tener en cuenta la autocorrelación y la verdadera varianza del proceso debe ser estimada. [Montgomery and Mastrangelo, 1991].

Montgomery and Mastrangelo [1991], proponen una carta EWMA que tiene en cuenta la presencia de datos autocorrelacionados en un proceso de control estadístico. La carta es construida asumiendo que los datos tienen una estructura de autocorrelación del tipo $ARIMA(0,1,1)$, para luego aplicarlo a la estadística EWMA, con el fin de obtener límites de control apropiados. Los autores mencionados aseguran

que si las observaciones del proceso son modeladas por un miembro de la familia ARIMA(p, d, q) y además están autocorrelacionadas positivamente y el proceso no cambia tan rápidamente en media, se puede usar la carta para datos autocorrelacionados con una elección apropiada de λ , donde λ es el peso dado a la media del subgrupo racional más reciente, es decir el parámetro de suavización del gráfico de control EWMA.

Alwan and Roberts [1988] construyen una carta basada en los residuales del modelo que mejor se ajuste a los datos y los lleva a las cartas tradicionales tipo Shewhart. Jiang and Tsui [2000] tienen en cuenta la estructura de autocorrelación de los datos en la construcción de los límites de las cartas de control, obteniendo límites fijos.

Dentro de las técnicas para lidiar con la autocorrelación se destaca la aplicación de un enfoque no paramétrico basado en técnicas de remuestreo. Entre las técnicas de remuestreo se encuentra el método bootstrap propuesto por Efron and Tibshirani [1986], que ha resultado ser una herramienta no paramétrica poderosa para aproximar las distribuciones muestral y varianza de estadísticos complicados para el caso cuando se tienen observaciones independientes idénticamente distribuidas (i.i.d.). Scepi [2002] proponen el diseño de un gráfico de control no paramétrico adoptando un algoritmo de remuestreo desarrollado por Politis and Romano [1994], junto con un método de tres vías para análisis de datos (STATIS, Escoufir, 1987) para controlar un proceso observado en sucesivos periodos de tiempo como para detectar unidades responsables de situaciones fuera de control.

Para el caso de dependencia entre observaciones la metodología propuesta por Bajgier [1992] y Liu and Tang [1996] consiste en la extracción de muestras con reemplazo de datos originales, cálculo de medias y localización de cuantiles para la construcción de límites de control. Chaparro and Vargas [2000]) presentan una metodología heurística para la construcción de gráficas de control para la media de un proceso que ajusta los límites de acuerdo a la dirección de la asimetría que presenta la función de la distribución de la característica de interés.

Sin suposiciones paramétricas Liu and Tang [1996] aplican y modifican el método de moving blocks bootstrap propuesto por Kunsch [1989] y Liu and Singh [1993]) para desarrollar gráficos de control válidos para datos débilmente dependientes. Sin embargo, la mayoría de métodos propuestos para datos multivariados son adecuados cuando los datos presentan una distribución normal multivariante.

1.2. Control estadístico de la calidad y metodología Seis Sigma

1.2.1. Calidad

La calidad se ha convertido en uno de los factores de decisión del consumidor más importantes en la selección de productos y servicios competidores. Esto se ha extendido, independientemente de si el consumidor es un individuo, una organización industrial, una tienda minorista, una institución financiera, o un programa de defensa militar. Consencuentemente, entender y mejorar la calidad es un factor clave para el éxito comercial, el crecimiento y la mejora de la competitividad. Se define calidad como adecuación para uso (calidad de diseño y calidad de conformidad). Ésta es inversamente proporcional a la variabilidad y el objetivo principal es intentar reducir la variabilidad en procesos y productos a través de la mejora de calidad. Este concepto se relaciona con una o más características deseables que debería poseer un producto o servicio. Es uno de los factores de decisión más importantes y por lo tanto entender y mejorar la calidad es un factor clave que lleva al éxito en los negocios, al crecimiento y a una posición competitiva fortalecida.

1.2.2. Control estadístico de procesos

Para diseñar los gráficos de control propuestos en la presente tesis, se hace uso de la aplicación de técnicas pertenecientes al control estadístico de la calidad (CEC) el cual se define como el uso de herramientas basadas en la estadística y técnicas principalmente diseñadas para la administración y mejora de procesos.

Una de las claves cruciales para comprender la medición del rendimiento, y por lo tanto, el control estadístico de procesos es la variación, se distingue entre dos causas de variación dentro de un proceso; aquellas que no pueden ser identificadas y corregidas que se denominan “causas fortuitas o no asignables”, y otras que pueden ser identificadas y se deben al proceso en si, denominadas “causas asignables”.

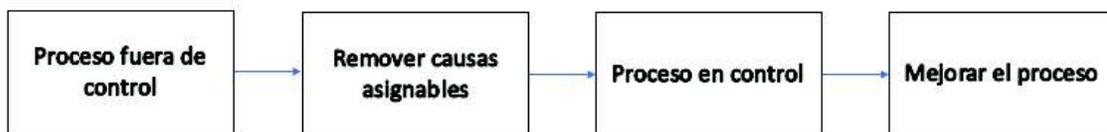
Proceso bajo control

Un proceso sujeto solamente a causas no asignables se denomina bajo control, esto es, las fluctuaciones aleatorias que esperamos en cualquier conjunto de mediciones. En la práctica, generalmente no existen procesos que se encuentren natural-

mente bajo control, causas especiales de variación se hacen presentes en el proceso, y cuando ocurren, cambian la variabilidad (aumento de la dispersión), descentrado (traslación del valor medio de la característica de calidad) o la forma de la distribución de las salidas del proceso.

En ocasiones, causas especiales de variación pueden ser resueltas localmente, en otros casos en el que no se pueda detener una reaparición, se podría tomar una acción para mitigar los efectos. Debido a esto, la clave es identificar rápidamente la ocurrencia de causas asignables que incidan en cambios en el proceso de tal forma que se tomen acciones correctivas antes que se fabriquen unidades no conformes.

En resumen, la primera tarea consiste en identificar y eliminar causas asignables de variación, a partir de aquello se mejora el proceso (Figura 1.4). Solo cuando el proceso esté en control se podrá predecir rendimiento futuro y capacidad del proceso. [Stapenhurst, 2013]



Fuente: Arman, H. Quality Management SPC-I. Recuperado de <http://slideplayer.com/slide/6149622/>

Figura 1.1: Proceso mejora proceso

1.2.3. Metodología Seis Sigma

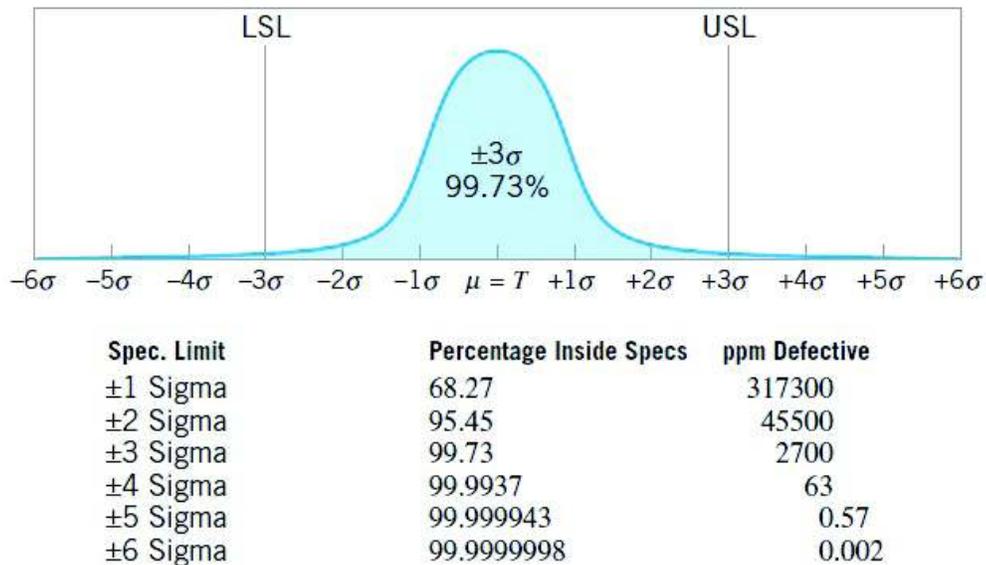
Seis Sigma es un enfoque revolucionario de gestión que mide y mejora la calidad, ha llegado a ser un método de referencia para satisfacer las necesidades del cliente y al mismo tiempo, lograrlo en niveles próximos a la perfección. [GestioPolis, 2001]. Dicho en pocas palabras, es una metodología de mejora de procesos que busca la reducción de la variación, defectos y otros tipos de errores para así, llegar a un máximo de 3.4 defectos por millón.

Este término fue acuñado por Motorola en los años 80 para denominar su iniciativa de reducción radical de defectos en productos. Renació, a mediados de los años 90, con un gran impulso, gracias a la acertividad con la que General Electric la aplicó en su organización (fabricación y servicios) y, sobre todo, por sus resultados económicos espectaculares.

La figura 1.2 muestra una distribución de probabilidad normal como modelo de una

característica de calidad con límites de especificación de tres desviaciones estándar en cada lado de la media. [Montgomery, 2007]

De acuerdo al gráfico, la probabilidad de producir un producto dentro de estas



Fuente: Montgomery [2007]

Figura 1.2: Variable normal centrada en $\mu = T$, porcentaje de elementos no conformes según el criterio Seis Sigma (capacidad de procesos). Se muestran los defectos por millón (ppm Defective) esperados para cada nivel de calidad del producto o capacidad de proceso.

especificaciones es de 0.9973, que corresponde a 2700 partes por millón (ppm). Esto se conoce como *rendimiento de calidad Tres Sigma*. Sin embargo, supongamos que consiste del ensamblaje de 100 componentes independientes las cuales deben ser no defectuosas para que el producto funcione satisfactoriamente. La probabilidad de que una unidad de producto específica sea no defectuosa es

$$0.9973 \times 0.9973 \times \dots \times 0.9973 = (0.9973)^{100} = 0.7631$$

Esto es, alrededor del 23.7% de productos elaborados bajo la calidad Tres Sigma serán defectuosos. Tomando en cuenta que muchos productos usados por la sociedad de hoy están compuestos de muchas partes, estos niveles de calidad son inaceptables. En la actualidad, Seis Sigma ha evolucionado desde su aplicación meramente como herramienta de calidad a incluirse dentro de los valores claves de algunas empresas, como parte de su filosofía de actuación.

Proceso DMAIC

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

- **(D)efinir** el sistema, los clientes y sus requisitos, además de los objetivos del proyecto.
- **(M)edir** los aspectos claves del proceso actual y recopilar datos relevantes; calcular la capacidad del proceso actual.
- **(A)nalizar** los datos para investigar y verificar las relaciones de causa y efecto..
- **(M)ejorar** u optimizar el proceso actual basándose en el análisis de datos utilizando técnicas como el diseño de experimentos, poka yoke o corrección de errores, y el trabajo estándar para crear un nuevo proceso de estado futuro.
- **(C)ontrolar** el proceso de estado futuro para garantizar que cualquier desviación del objetivo se corrija antes de que produzca defectos. Implementar sistemas de control tales como control estadístico de procesos supervisando continuamente el proceso. Este proceso se repite hasta que se obtiene el nivel de calidad deseado.



Fuente: Javier Castiñeira (2015). Six Sigma. Recuperado: <https://sigemeco.wordpress.com/tag/dmaic/>

Figura 1.3: Etapas y técnicas asociadas al proceso DMAIC

Así, la metodología Seis Sigma proporciona el boceto de aplicación de las técnicas del CEC en las diferentes etapas de procedimiento de reducción de variabilidad.

Etapas Definir

Esta etapa se caracteriza por la identificación de las oportunidades de proyecto para el cual se verificará o validará su potencia de avance legítimo. Un proyecto debe ser importante tanto para clientes como para el negocio de tal manera que se debe acordar la utilidad potencial del proyecto.

Etapas Medir

El propósito de la etapa de medición es evaluar y entender el estado actual del proceso, lo cual involucra recolectar datos concernientes a calidad, costo y rendimiento del ciclo; datos, que deberán ser suficientes para permitir un análisis exhaustivo y comprensivo del desempeño actual del proceso con respecto a las métricas.

Etapas Analizar

A través del análisis, tanto estadístico como cualitativo el objetivo se basa en usar los datos de la etapa de medición para empezar a determinar las relaciones causa-efecto en el proceso y comprender las diferentes fuentes de variabilidad. y comprender, en otras palabras, se busca determinar las causas potenciales de los defectos, problemas de calidad, quejas de los clientes, inconvenientes en el rendimiento, o gastos e ineficiencias que motivaron el problema.

Etapas Mejorar

En base a la etapa de análisis, el objetivo es mejorar el problema buscando una solución, priorizar potenciales ideas según los requerimientos del cliente, es decir, busca dar solución a la problemática.

Etapas Controlar

El objetivo de la etapa de control es el completar y monitorear los resultados obtenidos en las etapas anteriores, así como documentar las nuevas especificaciones del proceso en estudio.

Las herramientas aplicadas en esta etapa se basan en la detección de errores y su corrección, siendo el gráfico de control la más representativa.

1.3. Gráficos de control

El gráfico de control es la herramienta más usada en el análisis de variaciones con el fin de evaluar, monitorear y controlar el rendimiento de los procesos de producción o servicios; mediante el uso de los gráficos de control se pretende identificar variaciones en el proceso debido a causas comunes y causas especiales.

A través de los años, los gráficos de control se han posicionado entre los medios de control operativo más importantes, hasta adquirir tanta importancia como los controles de costos y materiales. La tecnología moderna de los ordenadores ha facilitado la implementación de los gráficos de control para cualquier tipo de proceso, ya que es posible obtener y analizar los datos en tiempo real y en línea en el centro de trabajo. [Scepi and Acconcia, 1996]

Un gráfico de control es un esquema en el que se hace corresponder un punto a cada uno de los valores de un estadístico calculados sobre muestras sucesivas en general del mismo tamaño, tomadas durante la producción. El gráfico contiene una línea central que representa el valor promedio de la característica de calidad correspondiente al estado "en control", y dos líneas horizontales, denominadas *límite de control superior* (LCS) y *límite de control inferior* (LCI). Estos límites de control son elegidos de tal forma que si el proceso está en control, casi todos los puntos de la muestra caerán entre ellos, mientras los puntos caigan dentro de los límites de control se asume que el proceso está en control y no es necesaria una acción correctiva. Sin embargo, un punto que cae fuera se interpreta como evidencia que el proceso está fuera de control, e investigación y acciones correctivas son requeridas para buscar y eliminar la causa asignable o causas responsables por este comportamiento.

Entre las varias razones para el uso de los gráficos de control, Montgomery [2007] destaca las siguientes:

- **Son una técnica comprobada para reducir la variabilidad.** Un gráfico de control exitoso reducirá el desperdicio y la repetición de trabajo, que son los principales factores que reducen la productividad en cualquier operación. Si se reduce los desperdicios y las modificaciones, la productividad y capacidad de producción aumentan y los costos disminuyen.

- **Son efectivos en la prevención de defectos.** El gráfico de control ayuda a mantener el proceso bajo control, lo cual es consistente con la filosofía “Hazlo bien a la primera”. Nunca es más barato clasificar las unidades “buenas” de las unidades “malas” más adelante que construirlas correctamente al principio. Si no tiene un control efectivo del proceso, le está pagando a alguien para que fabrique un producto no conforme.
- **Previene ajustes innecesarios en el proceso.** Un gráfico de control puede distinguir entre ruido de fondo y variación anormal; ningún otro dispositivo, incluido un operador humano, es tan efectivo para hacer esta distinción. Si los operadores del proceso ajustan el proceso basándose en pruebas periódicas no relacionadas con un programa de control, a menudo reaccionarán de forma exagerada ante el ruido de fondo y realizarán ajustes innecesarios. Tales ajustes innecesarios en realidad pueden resultar en un deterioro del rendimiento del proceso.
- **Proveen información de diagnóstico.** Con frecuencia, el patrón de puntos en el gráfico de control contendrá información de valor de diagnóstico para un operador o ingeniero experimentado. Esta información permite la implementación de un cambio en el proceso que mejora su desempeño.
- **Proveen información acerca de la capacidad del proceso.** El gráfico de control proporciona información sobre el valor de los parámetros de proceso importantes y su estabilidad en el tiempo. Esto permite que se realice una estimación de la capacidad del proceso.

1.3.1. Gráficos de control y pruebas de hipótesis

El marco de pruebas de hipótesis resulta útil para analizar el rendimiento de un gráfico de control. La decisión de rechazar la hipótesis nula cuando esta es verdadera se denomina error tipo I, es decir, concluir que el proceso está fuera de control cuando está realmente en control. Por otro lado, cuando no se rechaza la hipótesis nula cuando esta es falsa se denomina error tipo II, es decir, concluir que el proceso está en control cuando en realidad está fuera de control. En la práctica hallamos estos cuatro escenarios:

Cuadro 1.1: Tipos de errores estadísticos

Situación Real:		
	H_0 es verdadera	H_a es verdadera
Aceptar H_0	Decisión correcta Prob= $1-\alpha$	Error Tipo II Prob= β
Rechazar H_0	Error Tipo I Prob= α (nivel de significancia)	Decisión correcta Prob= $1-\beta$ (poder de la prueba)

Para el monitoreo de una característica de calidad X , se observa una submuestra $S = \{X_1, X_2, \dots, X_n\}$ para cada instante de tiempo, la cual sigue una distribución F con media μ_0 y desviación estándar σ . Entonces, en cada instante de tiempo para evaluar si un proceso está bajo control, se prueba la hipótesis $H_0 : \mu = \mu_0$ contra $H_a : \mu \neq \mu_0$. Por tanto, se tiene que el proceso está fuera de control, si se rechaza H_0 , de otra manera, se dice que está bajo control. Este contraste, se puede realizar de forma equivalente mediante un gráfico de control que representa una secuencia de regiones de aceptación en el tiempo.

Un procedimiento de control consiste generalmente de dos fases. La primera se basa en la construcción del esquema de control, la definición de los límites de control y como consecuencia la identificación de la región de control. Muestras de observaciones son seleccionadas aleatoriamente del proceso, y medidas de posición y/o variabilidad son calculadas de cada muestra. El objetivo es calcular la tolerancia natural del proceso y su variabilidad natural.

Para definir límites robustos es importante la elección de estadísticos no tan sensibles a la presencia de puntos atípicos en la muestra. Al estar estos definidos, nuevas medidas son recogidas para dar comienzo a la fase de monitoreo en línea y nuevas muestras son dibujadas en el gráfico de control. Un punto muestral cayendo fuera de la región de control es interpretado como evidencia que el proceso está fuera de control. Por tanto, para controlar un proceso, es necesario usar estadísticos capaces de detectar cambios en el mismo.

Debido a que es una prueba estadística, es posible que el proceso se declare fuera de control cuando en realidad no lo está. En este caso, podemos decir que tenemos una "falsa alarma" α . En la fase I, es importante definir límites de control de tal manera que la tasa de falsa alarma sea pequeña. Fijado α es importante que los límites sean escogidos tanto robustos hacia atípicos como sensibles a la presencia de situaciones reales fuera de control.

El cálculo de los límites de control es muy sencillo, debido a que los parámetros de control (media o varianza) ya están definidos y para el cálculo de los límites se asume una distribución conocida para la característica de calidad, por ejemplo una distribución normal.

Sea w un estadístico de muestra que mide una característica de calidad de interés, y supongamos que la media de w es μ_w y la desviación estándar de w es σ_w . Entonces podemos definir un modelo general para los límites de control de un gráfico de control como:

$$LCS = \mu_w + L\sigma_w$$

$$LC = \mu_w$$

$$LCI = \mu_w - L\sigma_w$$

Donde L es la distancia de los límites de control respecto a la línea central expresada en unidades de desviación estándar.

Para el diseño de un gráfico de control hay que especificar el tamaño de muestra y la frecuencia de muestreo. En general, muestras más grandes facilitarán la detección de pequeños cambios en el proceso, sin embargo esto no es plausible debido al coste económico que representa, una de las formas de evaluar las decisiones con respecto al tamaño de muestra y frecuencia de muestreo es a través de la *longitud media de corrida* (ARL) que es el número promedio de puntos que deben ser graficados antes de que se indique una condición fuera de control. Si las observaciones en el proceso no están correlacionadas, el ARL se puede calcular para cualquier gráfico de control como:

$$ARL = \frac{1}{p}$$

donde p es la probabilidad de que cualquier punto supere los límites de control. Mediante esta ecuación se evalúa el rendimiento de los gráficos de control, aunque a veces es conveniente expresar este rendimiento en términos de su *Tiempo Promedio a la Señal* (ATS), que se define como

$$ATS = ARL \times h$$

donde h es el intervalo de tiempo fijo en horas, en el que se toman muestras. Cuanto menor sea el ARL de un gráfico de control mejor es su desempeño pues es capaz de detectar de manera más rápida un desajuste.

Para la medición de la potencia de la prueba en un contraste de hipótesis se utiliza la curva de potencia, el inverso de esta curva denominada *curva característica de ope-*

ración representa una visualización del error tipo II en función de distintos valores para la hipótesis alternativa. Dicha curva se construye representando este error en función de la magnitud del cambio que se pretende examinar.

1.3.2. Fase I y Fase II de un gráfico de control

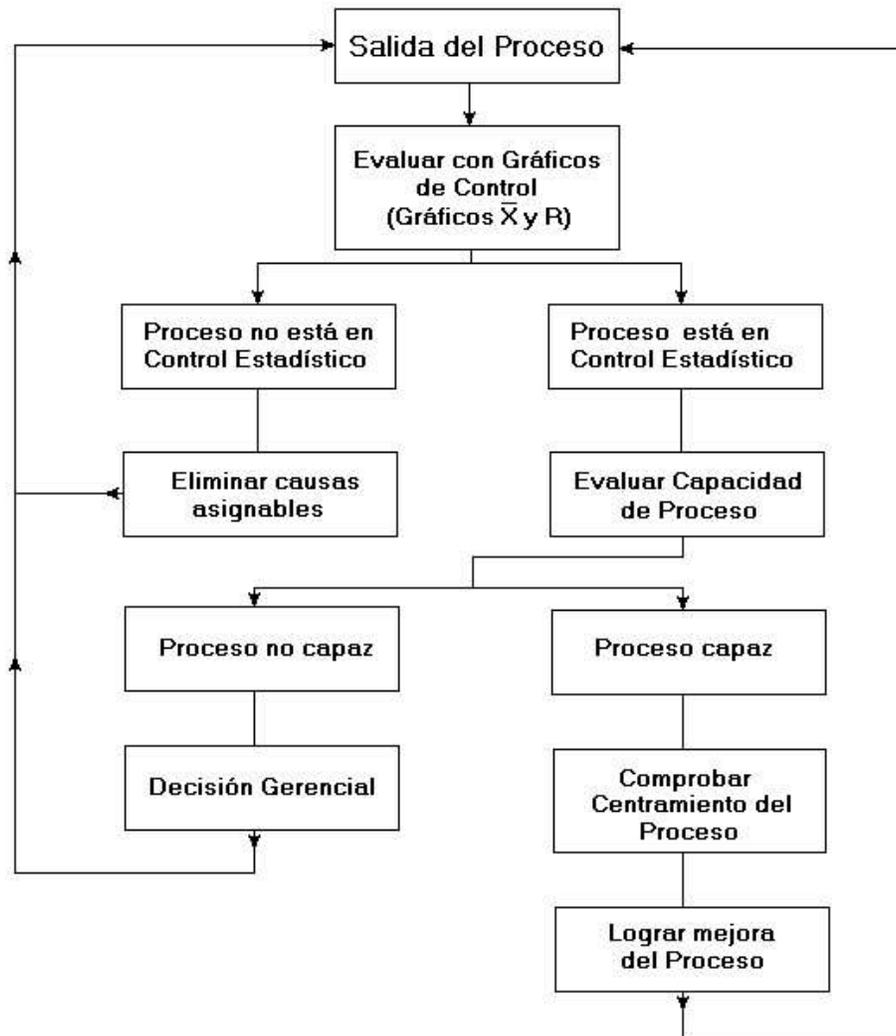
El uso del gráfico de control estándar implica aplicaciones de fase I y fase II, con dos objetivos diferentes y distintos. En la fase I, un conjunto de datos históricos procesados es recolectado y analizado a la vez en un análisis retrospectivo, usando esta muestra preliminar se estiman los límites de control. En esta fase se identifican las principales causas de variación asignables y se toman las medidas para corregirlas, hasta lograr la estimación de los límites de control de la variable CTQ correspondiente al proceso bajo control. Los límites de control se calculan como un intervalo de confianza. Se suelen tomar aquellos que distan de la media en tres desviaciones típicas ($\mu \pm 3\sigma$). Al conjunto de datos históricos se lo denomina muestra de calibrado, siendo el contraste de hipótesis (H_0) sobre una misma población F .

Así, en la fase I comparamos una colección de m puntos contra un conjunto de límites de control calculados a partir de esos puntos. Se recomienda que los límites de control naturales se estimen a partir de una muestra de calibrado constituida por 20 o más observaciones. [Montgomery, 2007]

Es bastante usual suponer que en la fase I el proceso está inicialmente fuera de control, por lo que el objetivo del analista es llevar el proceso a un estado de control estadístico. Los límites de control se calculan en función de los m subgrupos y los datos graficados en los gráficos de control. Se investigan los puntos que están fuera de los límites de control, en busca de posibles causas asignables. Los puntos fuera de los límites de control se excluyen y se calcula un nuevo conjunto de límites de control revisados. Luego se recopilan nuevos datos y se comparan con estos límites revisados. En ocasiones, este tipo de análisis requerirá varios ciclos en los que se emplee el gráfico de control, se detecten y corrijan las causas asignables, se calculen los límites de control revisados y se actualize y amplíe el plan de acción fuera de control.

Los tipos de causas asignables que usualmente ocurren en la fase I resultan en cambios de proceso bastante grandes. La longitud de corrida promedio (ARL) no es un buen indicador de rendimiento para esta fase, dado que se busca la probabilidad de que se detecte una causa asignable en lugar de la aparición de falsas alarmas.

En la fase II, usamos el gráfico de control para monitorear el proceso comparando el estadístico muestral para cada muestra sucesiva a medida que se extrae del proceso a los límites de control. Cuando las observaciones individuales de la variable CTQ de estudio X están dentro de los límites de control, se dice que el proceso está bajo control. El contraste que se realiza es si una nueva observación de la población G pertenece a la población F . En este caso la longitud de corrida promedio (ARL) es una base válida para evaluar el rendimiento de un gráfico de control en la fase II.



Fuente: Ivan Escalona. Gráficos de Control de Shewart. Recuperado: <http://www.monografias.com/trabajos12/concalgra/concalgra.shtml>

Figura 1.4: Proceso de monitoreo de una variable que afecta la calidad de un producto mediante la aplicación de gráficos de control

1.3.3. Tipos de gráficos de control

Gráficos de control univariantes

Los gráficos de control se pueden dividir en dos categorías. Si la característica de calidad puede ser medida y expresada como un número en una escala continua de medición, es usualmente denominado **variable**. En tal caso, es conveniente describir la característica de calidad con una medida de tendencia central y una medida de variabilidad, gráficos de control para tendencia central y variabilidad son llamados **gráficos de control por variables**.

Muchas características de calidad no son medidas en una escala continua o incluso una escala cuantitativa, podemos juzgar cada unidad de producto como conforme o no conforme sobre la base de si posee o no ciertas atributos, o podemos contar el número de no conformidades (defectos) que aparecen en una unidad de producto, de ser este el caso, podemos denominar a estos gráficos de control **gráficos de control por atributos**.

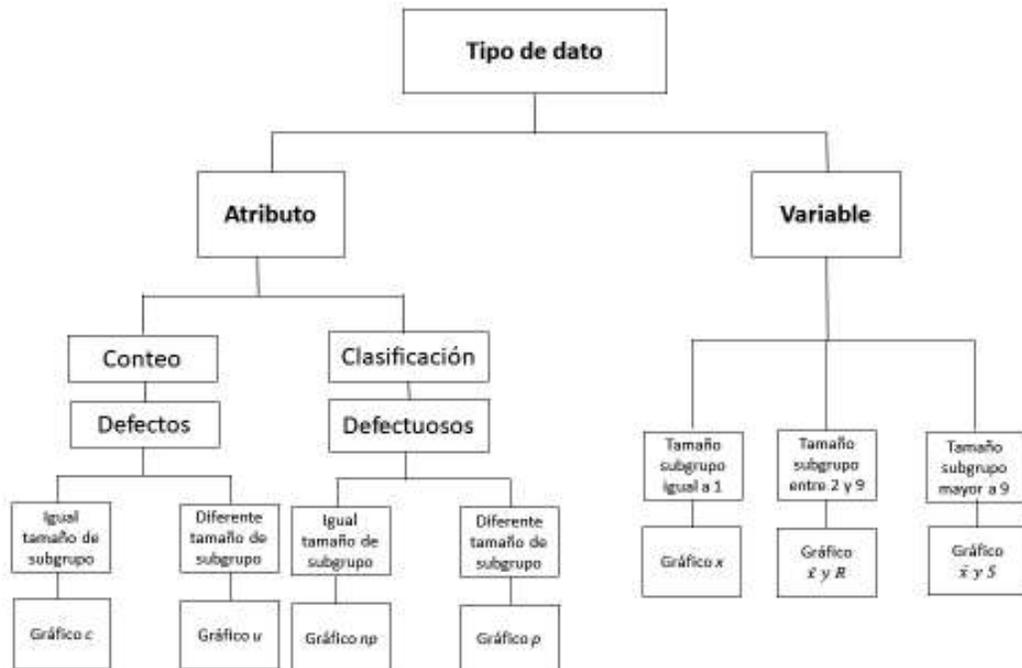


Figura 1.5: Clasificación de gráficos de control de acuerdo a la variable CTQ

La figura 1.5 muestra los gráficos de control univariantes más comunes, así como el tipo de observación sobre la cual se recomienda su uso, cabe recalcar que la aplicación de estos gráficos de control se basa en independencia y en una distribución

paramétrica en los datos, por ejemplo para los gráficos por atributos se basa en una distribución de poisson y una distribución binomial para gráficos para defectos y defectuosos respectivamente, para el caso de gráficos por variables se basa en que los datos están distribuidos bajo una distribución normal.

Gráficos de control para \bar{X} y R.

Cuando una característica de calidad es una variable, es necesario controlar tanto el valor medio de la característica como su variabilidad. El gráfico \bar{x} se usa para controlar la media del proceso, mientras que la variabilidad se puede monitorear con un gráfico de control para el rango denominado gráfico de control R.

Para construir límites de control, debemos calcular la media y el rango de cada muestra \bar{x}_i y R_i , respectivamente $i = 1, \dots, n$. La media global y el rango se calcularán como:

$$\bar{R} = \frac{\sum_i R_i}{k} \text{ y } \bar{\bar{x}} = \frac{\sum_i \bar{x}_i}{k}$$

Puesto que R_i/d_2 es un estimador insesgado, donde d_2 es un estadístico de rango que representa la media del rango relativo dado por $W = \frac{R}{\sigma}$, luego, $\frac{\bar{R}}{d_2} = \frac{\sum_i R_i}{d_2 k}$ se usará como estimador centrado de la desviación estándar teórica. A partir de eso hay que contrastar si cada valor \bar{x}_i está en el intervalo

$$\left[\bar{\bar{x}} - \frac{3\bar{R}}{\sqrt{nd_2}}, \bar{\bar{x}} + \frac{3\bar{R}}{\sqrt{nd_2}} \right]$$

con una probabilidad aproximada igual al 97,73 % (usando intervalos de amplitud 3σ).

Entonces los límites de control son:

$$LCI = \bar{\bar{x}} - \frac{3}{\sqrt{nd_2}}\bar{R} = \bar{\bar{x}} - A_2\bar{R}$$

$$LC = \bar{\bar{x}}$$

$$LCS = \bar{\bar{x}} + \frac{3}{\sqrt{nd_2}}\bar{R} = \bar{\bar{x}} + A_2\bar{R}$$

donde A_2 es una constante tabulada para varios tamaños de muestra.

Se ilustrarán los gráficos de control usando el paquete **qcr** en R para control estadístico de calidad. El paquete incluye un conjunto completo de herramientas del CEC univariadas y multivariadas que completa e incrementa las técnicas de CEC disponibles en R. Combina procedimientos SQC flexibles, tradicionales y nuevos para tratar problemas reales de control de calidad en la industria y la consultoría. La función `qcs.xbar()` incluida en el paquete **qcr** proporciona una manera simple

de calcular la línea central y los límites de control utilizando el rango como desviación estándar para el gráfico.

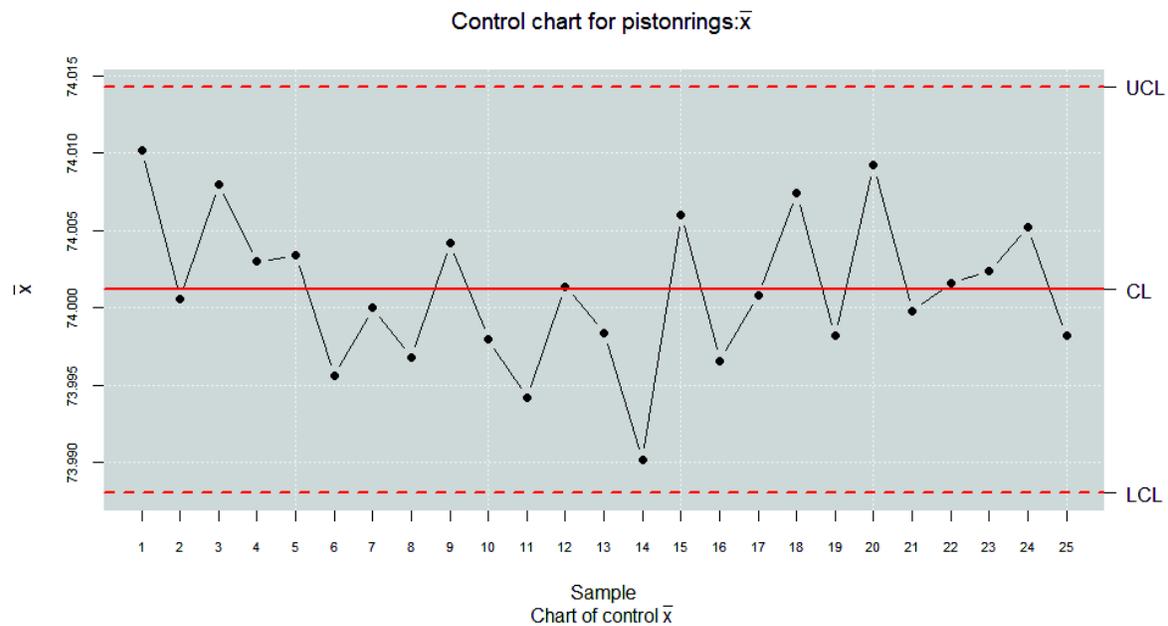


Figura 1.6: Gráfico de control \bar{x} usando el rango como medida de variabilidad

La variabilidad del proceso en un gráfico de control puede monitorearse trazando valores del rango R . Para un tamaño de muestra n dado, la línea central y los límites de control del gráfico R son los siguientes:

$$\begin{aligned}
 LCI &= D_3 \bar{R} \\
 LC &= \bar{R} \\
 LCS &= D_4 \bar{R}
 \end{aligned}$$

donde D_4 y D_3 son constantes tabuladas para varios valores de n .

Estos límites de control así como también la línea central se calculan usando la función $qcs.R()$. Debido a que los límites de control en el gráfico \bar{x} dependen de la variabilidad del proceso, es mejor comenzar con el gráfico R . El gráfico R se muestra en la figura 1.7 mientras que el gráfico \bar{x} se traza en la figura 1.6. No se observa ninguna indicación de un punto fuera de los límites de control, por lo tanto, el proceso está bajo control.

Gráficos de control para \bar{X} y S .

A veces es deseable estimar la desviación estándar del proceso directamente en lugar de indirectamente mediante el uso del rango R , esto lleva a gráficos de control

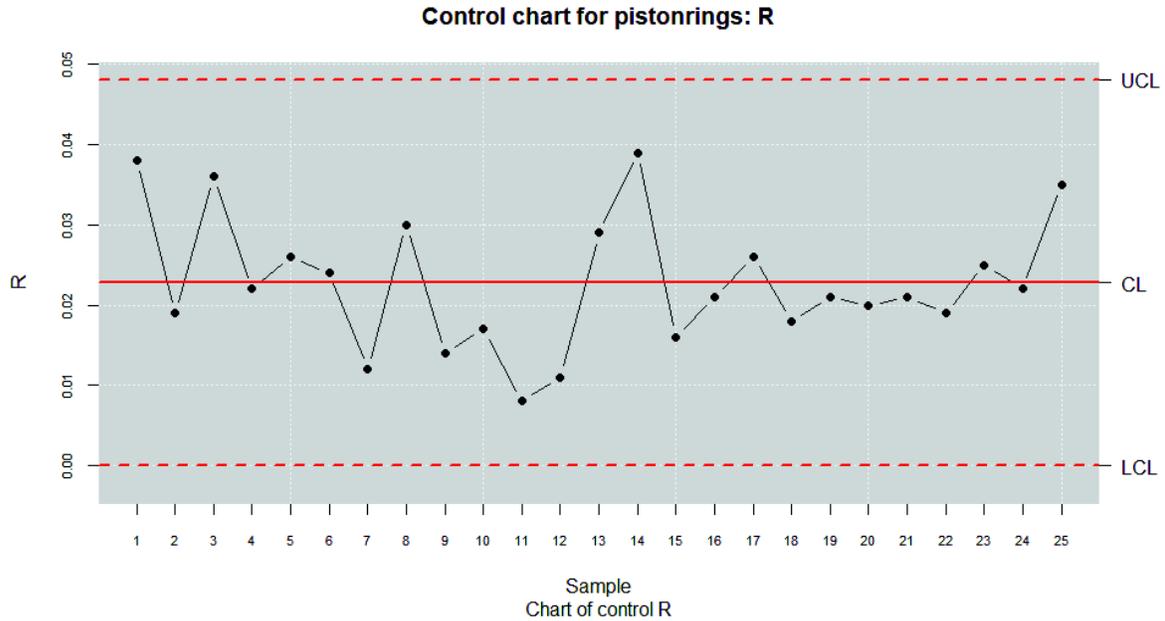


Figura 1.7: Gráfico de control R

para \bar{x} y S , donde s es la desviación estándar de la muestra.

Para la construcción de los límites de control, la media y la desviación deben calcularse para cada muestra \bar{x}_i y s_i , respectivamente $i = 1, \dots, n$. La media global es un estimador insesgado de la media teórica, mientras que la desviación estándar muestral es un estimador sesgado de la desviación estándar teórica, luego

$$\hat{\sigma} = \frac{\bar{s}}{c_2} = \frac{\sum_i s_i}{c_2 k}$$

se usará como un estimador centrado de la desviación estándar teórica, donde c_2 es una constante que depende del tamaño de muestra n .

Cada valor \bar{x}_i debe estar en el intervalo

$$\left[\bar{\bar{x}} - \frac{3\bar{s}}{\sqrt{nc_2}}, \bar{\bar{x}} + \frac{3\bar{s}}{\sqrt{nc_2}} \right]$$

con una probabilidad aproximada igual al 97,73 % (usando intervalos de amplitud 3σ).

Entonces, los límites de control para la media son:

$$\begin{aligned} LCI &= \bar{\bar{x}} - \frac{3}{\sqrt{nc_2}} \bar{s} = \bar{\bar{x}} - A_1 \bar{s} \\ LC &= \bar{\bar{x}} \\ LCS &= \bar{\bar{x}} + \frac{3}{\sqrt{nc_2}} \bar{s} = \bar{\bar{x}} + A_1 \bar{s} \end{aligned}$$

Los límites de control se calculan de manera similar a como en el gráfico anterior,

la función `qcs.xbar()` se usa para determinar estos límites, siendo la diferencia la elección de la desviación estándar.

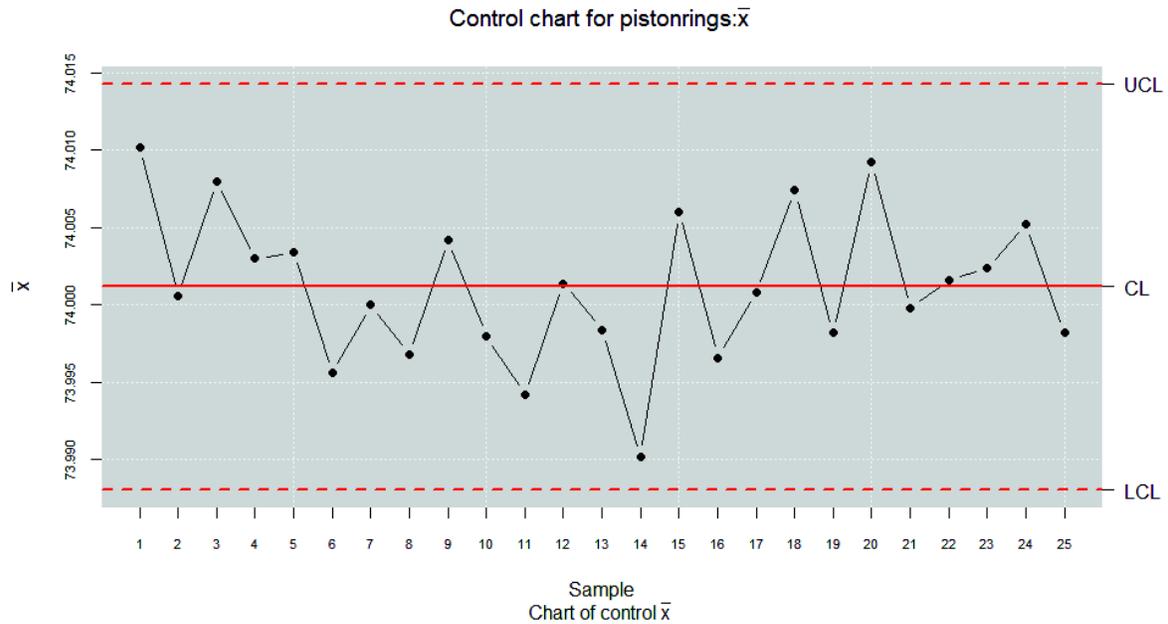


Figura 1.8: Gráfico de control \bar{x} usando la desviación estándar como medida de variabilidad

La desviación estándar se debe estimar mediante el análisis de datos pasados. Supongamos que hay disponibles m muestras preliminares, cada una de tamaño n , y sea s_i la desviación estándar de la i -ésima muestra. El promedio de las desviaciones estándar de m es $\bar{s} = \frac{1}{m} \sum_{i=1}^m s_i$.

El estadístico \bar{s}/c_4 es un estimador insesgado de σ . Entonces la línea central y los límites de control son los siguientes:

$$LCI = B_3\bar{s}$$

$$LC = \bar{s}$$

$$LCS = B_4\bar{s}$$

donde A_1 , B_3 y B_4 son constantes tabuladas para varios valores de n .

La función `qcs.S()` calcula los límites de control así como también la línea central del gráfico. Los gráficos de control para \bar{x} y s se muestran en las figuras 1.8 y 1.9, respectivamente, y no hay ninguna indicación de que el proceso esté fuera de control.

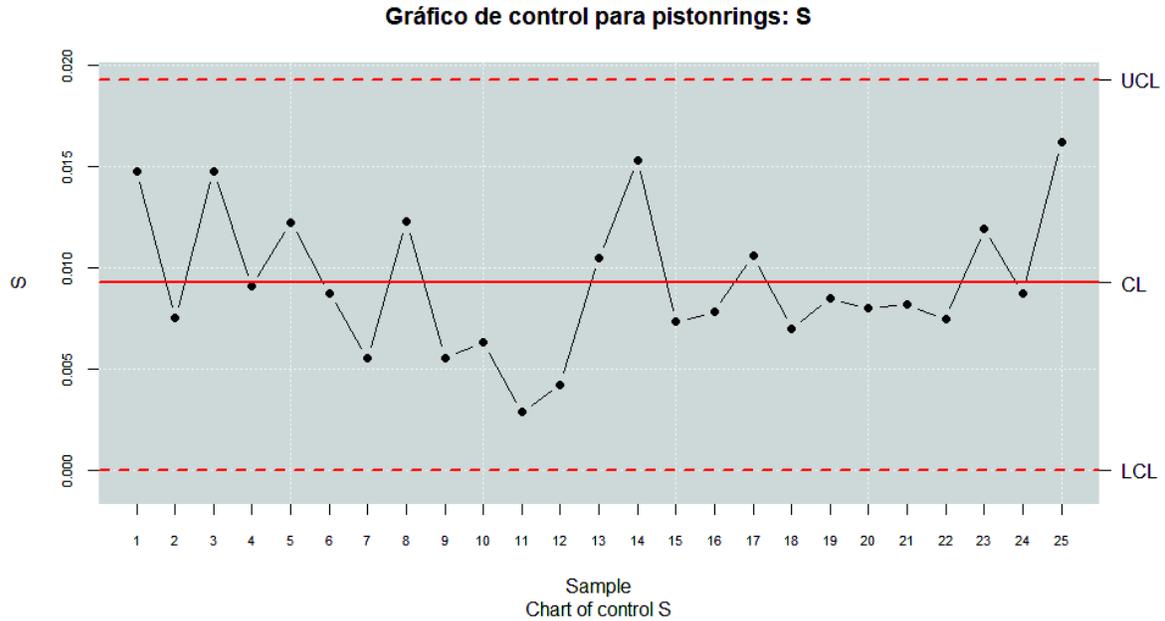


Figura 1.9: Gráfico de control S

Gráfico CUSUM

Gráficos de control de suma acumulativa fueron propuestos por Page [1961] para monitorear la media del proceso, se pueden construir sumas acumuladas tanto para observaciones individuales como para los promedios de subgrupos racionales. El estadístico s_r se usará para determinar si el proceso está bajo control o no.

$$s_r = \sum_{i=1}^r (\bar{x}_i - \mu_0) \sim N \left(r(\mu - \mu_0), \frac{r\sigma_0^2}{n} \right)$$

bajo supuestos de normalidad, donde μ_0 es el objetivo para el promedio del proceso.

Gráfico de control EWMA

El gráfico de control del promedio móvil ponderado exponencialmente (EWMA) fue presentada por Roberts [1959] como una alternativa cuando estamos interesados en detectar pequeños cambios. El rendimiento del gráfico de control de EWMA es aproximadamente equivalente al del gráfico de control de suma acumulada, ya que ambos supervisan el promedio del proceso.

Los promedios contemporáneos y de observaciones están representados en el gráfico, donde se da más peso a las observaciones más recientes (promedios móviles y_i),

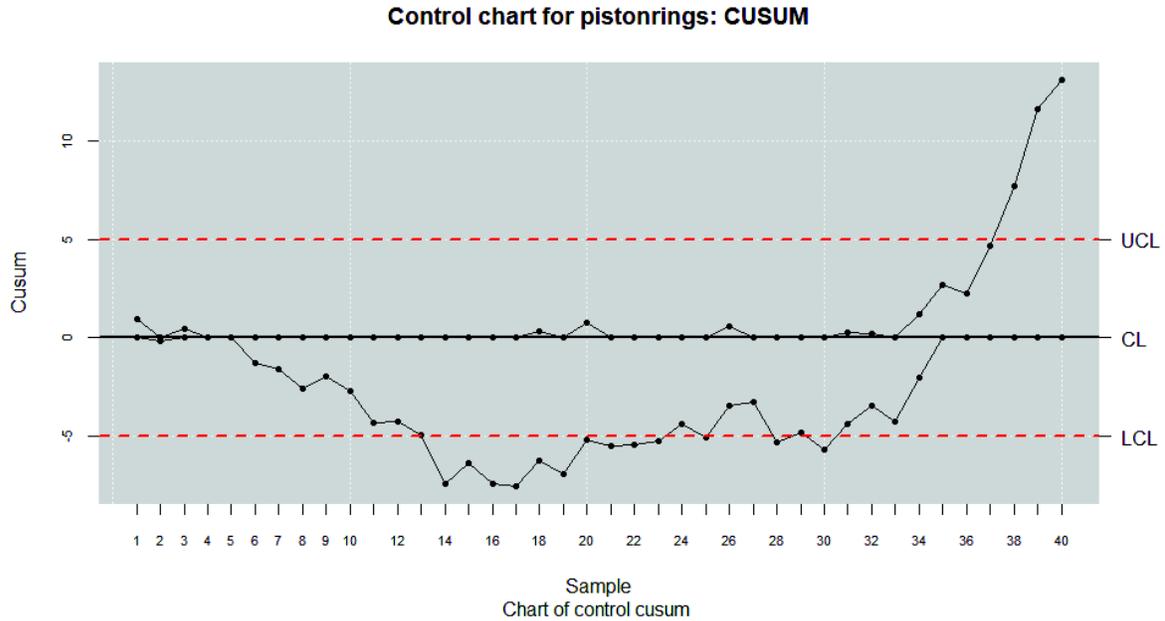


Figura 1.10: Gráfico de control *CUSUM*

y se define como

$$y_i = \lambda x_i + (1 - \lambda)y_{i-1}$$

donde $0 < \lambda \leq 1$ y el valor inicial ($y_0 = \mu_0$) es el objetivo del proceso.

Por lo tanto, el gráfico de control EWMA se construiría trazando y_i frente al número de muestra i (o tiempo). La línea central y los límites de control para la tabla de control EWMA son los siguientes.

$$LCI = \mu_0 - 3\sigma \sqrt{\frac{\lambda(1-(1-\lambda)^{2i})}{2-\lambda}}$$

$$LC = \mu_0$$

$$LCS = \mu_0 + 3\sigma \sqrt{\frac{\lambda(1-(1-\lambda)^{2i})}{2-\lambda}}$$

Si el valor de i es alto, los límites de control son:

$$LCI = \mu_0 - 3\sigma \sqrt{\frac{\lambda}{2-\lambda}}$$

$$LC = \mu_0$$

$$LCS = \mu_0 + 3\sigma \sqrt{\frac{\lambda}{2-\lambda}}$$

El gráfico de control se traza de la misma manera que el gráfico *CUSUM* para el cálculo de los estadísticos requeridos.

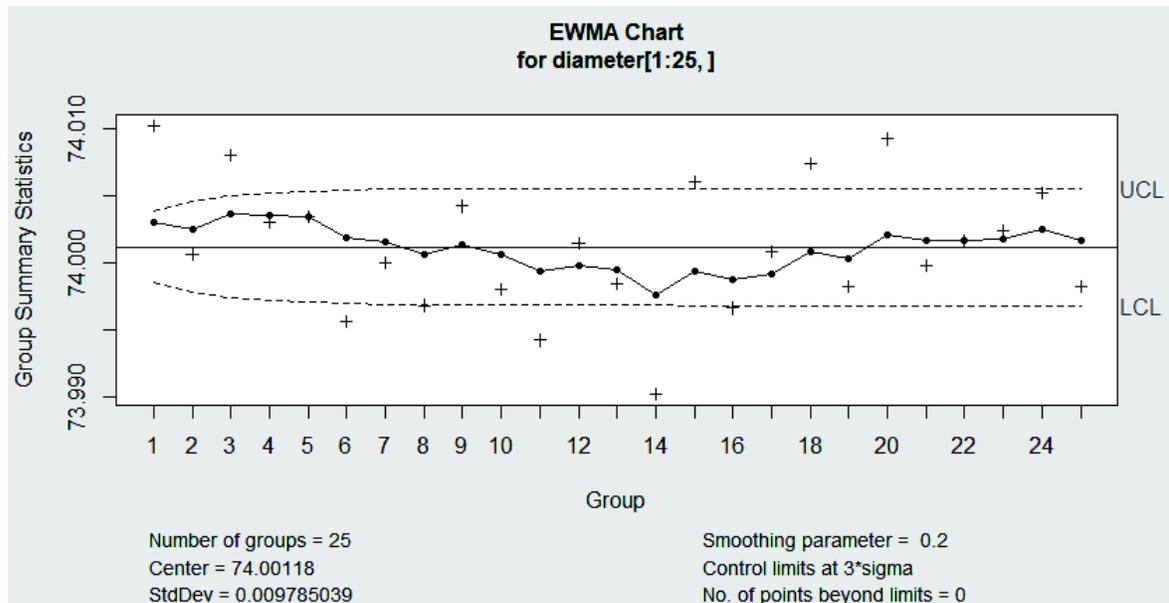


Figura 1.11: Gráfico de control *EWMA*

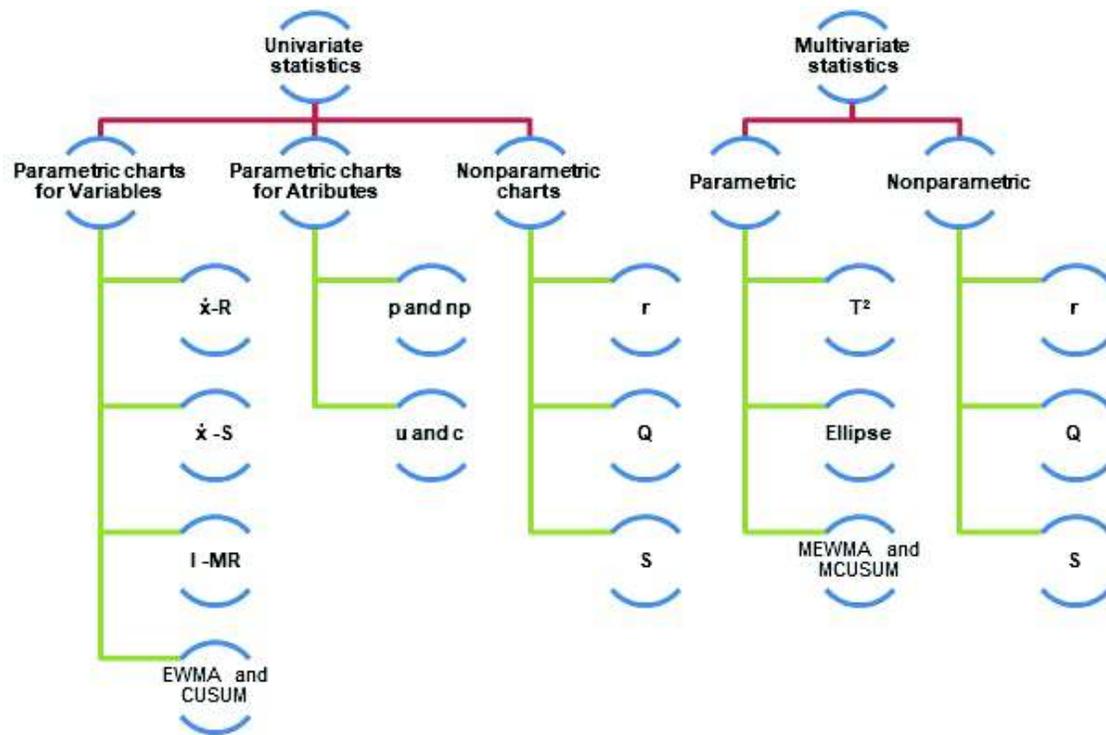
Gráficos de control multivariantes

Cuando varias variables aleatorias caracterizan la calidad de un proceso/servicio, es necesario aplicar técnicas estadísticas de control de calidad multivariante. Existen dos retos formidables en el análisis de procesos de datos multivariantes. Primero, los datos se vuelven más dispersos mientras aumenta la dimensión. Requerimientos de tamaños de muestras son formidables y una muestra de referencia de cientos de miles de observaciones puede ser necesitada para caracterizar por completo un proceso en control si tres o más variables de calidad (que no siguen una distribución normal multivariante) son medidas. Usar gráficos de control individuales para monitorear las p características independientes separadamente puede ser muy engañoso. Si una de las variables está fuera de control, el sistema resultaría fuera de control. Si $P(\text{Error Tipo I}) = \alpha$, entonces el error Tipo I para el control conjunto de las p medidas independientes, puede ser muy grande para un valor grande de p , $\alpha' = 1 - (1 - \alpha)^p$. De igual forma, si todas las variables están fuera de control, todo el sistema estaría fuera de control. Si $P(\text{Error Tipo II}) = \beta$, entonces el error Tipo II para el control conjunto de las p medidas independientes, puede ser muy grande para un valor grande de p , $\beta' = 1 - (1 - \beta)^p$.

Por lo tanto, los límites de control son diferentes a los construidos; si las variables son dependientes, el cálculo del nivel de significancia, α , se vuelve complicado. Este tema es particularmente importante hoy en día, ya que los procedimientos de inspección automáticos hacen que sea relativamente fácil medir muchos parámetros

en cada unidad de producto fabricado.[Montgomery, 2007]

Más aún, existen cambios en la localización o escala mientras no sea posible controlar la estructura de correlación entre dos (o más) variables. En este sentido, es posible tener dos variables las cuales estén ambas en control cuando son consideradas por separado en un sentido univariante, pero fuera de control cuando son consideradas en conjunto bajo un esquema multivariante [Wierda, 1994].



Fuente: Miguel Flores (2018). Statistical Quality Control with qcr package.

Figura 1.12: Clasificación de gráficos de control de acuerdo al tipo de muestra y distribución de las observaciones

En la figura 1.12 se muestran los distintos tipos de gráficos de control estándar dependiendo si el proceso está definido para uno o más variables y de si se asume o no una distribución paramétrica para las observaciones.

Capítulo 2

Marco teórico

2.1. Profundidad de datos

2.1.1. Introducción

La noción de profundidad de datos fue propuesta por Tukey (1975) como una herramienta gráfica para visualizar conjuntos de datos bivariantes, la cual se basa en una mediana multivariante que es el punto más “profundo” en una nube de puntos dada en \mathbb{R}^d .

Al medir la profundidad de un punto arbitrario z con respecto a los datos, Donoho y Gasko (1992) consideraron hiperplanos a través de z y determinaron su profundidad por la porción más pequeña de datos que están separados por dicho hiperplano. Desde entonces, esta idea ha resultado extremadamente fructífera. [Mosler, 2013]. Nociones generales de profundidad de datos se han introducido al igual que especiales, variando de acuerdo a su computabilidad, robustez y su sensibilidad a reflejar formas asimétricas en las observaciones.

Los conjuntos de nivel superior de un estadístico de profundidad proporcionan una familia de estadísticos denominados *regiones recortadas en profundidad* o *regiones centrales*, estas describen la distribución respecto a su ubicación, escala y forma, con la región central sirviendo como *mediana*.

La noción de profundidad se ha extendido desde nubes de puntos, que son distribuciones empíricas, a distribuciones probabilísticas generales en \mathbb{R}^d , en consecuencia permitiendo la ley de los grandes números y resultados de consistencia. También se ha extendido desde datos multivariantes hasta datos en espacios funcionales.

2.1.2. Conceptos básicos

La profundidad de datos se basa en el grado de centralidad de un punto dado respecto a una distribución de probabilidad, este concepto proporciona un ordenamiento de puntos de adentro hacia afuera en cualquier dimensión y conduce a un acercamiento estadístico no-paramétrico multivariante.

A partir de una noción de profundidad y de una distribución de probabilidad, podemos definir un *orden* en el espacio euclideano multidimensional. Este sería un orden respecto al grado de centralidad respecto de la distribución, en concreto, un orden según el grado de centralidad no será un orden total y su interpretación no es tan inmediata como la relación de orden en \mathbb{R} , pero se puede extender las técnicas de datos univariantes basadas en el orden natural al campo multivariante.[Casco et al., 2011]

Antes de la popularización del concepto de profundidad estadística, Barnett [1976] considera el orden inducido por el *convex hull peeling* el cual, dado una nube de puntos multivariantes y fijado $i = 1$, consiste en:

1. Trazar la envolvente convexa de los puntos de la nube. Al conjunto de vértices lo denotamos por C_i
2. Eliminar de la nube los puntos C_i
3. Incrementar i en una unidad y repetir 1 y 2 mientras queden puntos en la muestra.

La figura 2.1 muestra el procedimiento del *convex hull peeling* y así, los puntos más externos de la nube estarán en C_1 , y a medida que aumenta i , aumentará el grado de centralidad de los puntos de C_i . El conjunto de puntos más centrales será C_m , donde m es el mayor índice alcanzado.

Este procedimiento es muy conocido en el análisis exploratorio de datos, sin embargo, desde el punto de vista de la estadística tiene el enorme inconveniente de que no se dispone de ninguna representación poblacional del mismo. Es decir, dada una distribución de probabilidad genérica P , no se puede definir el *convex hull peeling*. Por tal motivo, los resultados asintóticos sobre el *convex hull peeling* son muy escasos y complejos, ya que los estadísticos no tienen un límite claro al cual converger. [Casco et al., 2011]

DEFINICIÓN 2.1. Sea F una distribución de probabilidad en \mathbb{R}^d , $d \geq 1$. Supongamos que F es absolutamente continua y que $\{X_1, \dots, X_n\}$ es una muestra aleatoria de F .

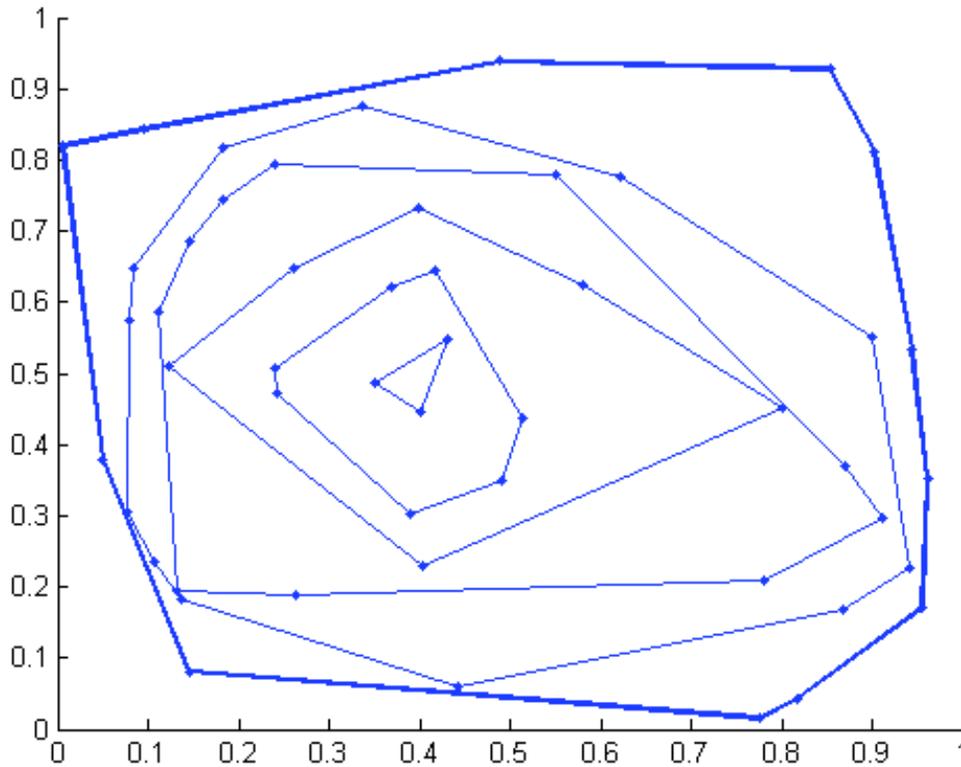


Figura 2.1: Ilustración procedimiento de convex hull peeling

Nótese que cada punto X_i es visto como un vector columna de dimensión $d \times 1$. Una función de profundidad es una función acotada $(\cdot, F) : \mathbb{R}^d \mapsto [0, 1]$ que asigna a cada punto $x \in \mathbb{R}^d$ su grado de centralidad respecto a F o respecto a una nube de puntos $\{X_1, \dots, X_n\}$

Idealmente debe satisfacer las siguientes propiedades: [Casco et al., 2011]

- (D₁) *Invarianza afín:* La profundidad de un punto $x \in \mathbb{R}^d$ no depende del sistema de coordenadas elegido, en particular de la escala del sistema de medición. Es decir, para cualquier matriz no singular $A \in \mathbb{R}^{d \times d}$ y $b \in \mathbb{R}^d$, se tiene $D(Ax + b, F_{AX+b}) = D(x, F_X)$, donde F_X denota la probabilidad inducida por el vector aleatorio.
- (D₂) *Profundidad máxima en el centro:* Para una distribución que tenga un único centro definido, la profundidad se maximiza en ese punto, es decir, $D(\theta, F) = \sup_{x \in \mathbb{R}^d} D(x, F)$ si θ es el centro de F .
- (D₃) *Monotonía respecto al punto de máxima profundidad:* A medida que $x \in \mathbb{R}^d$ se aleja del punto más profundo a lo largo de cualquier rayo fijo la profundidad de x

debería decrecer, es decir, si θ es un punto de máxima profundidad, entonces $D(x, F) \leq D(\theta + \lambda(x - \theta), F)$ para cualquier $0 \leq \lambda \leq 1$.

(D_4) *Se anula en el infinito*: La profundidad de un punto x debería aproximarse a cero si $\|x\|$ es demasiado grande, es decir,

$$\lim_{\|x\| \rightarrow \infty} D(x, F) = 0 \quad (2.1)$$

Existen medidas de profundidad que no satisfacen (D_4) y en que en lugar cumplen una condición de **cuasi-concavidad como función de x** , es decir que el conjunto $\{x : D(x, F) \geq c\}$ es convexo para cualquier número real c . [Fernández, 2013]

DEFINICIÓN 2.2. Sea $X \in \mathbb{R}^d$ un vector aleatorio entonces:

- X es *centralmente simétrico (C-simétrico)* respecto a θ si la distribución de $X - \theta$ es la misma que la de $\theta - X$.
- X es *angularmente simétrico (A-simétrico)* respecto a θ si $(X - \theta) / \|X - \theta\|$ es centralmente simétrico respecto al origen.
- X es *semiespacialmente simétrico (H-simétrico)* respecto a θ si $P(X \in H) \geq \frac{1}{2}$ para todo H semiespacio cerrado tal que $\theta \in H$.

2.1.3. Medidas de profundidad

Las funciones de profundidad para medidas multivariantes surgieron con el objetivo de extender la noción de los estadísticos de orden. Al ser una técnica no paramétrica resulta muy útil en el caso que la distribución de los datos no sea conocida. Se presentará algunos ejemplos de funciones de profundidad para datos multivariados y se clasificarán de acuerdo a las propiedades ($D_1 - D_4$).

Profundidad Semiespacial

Tukey definió la profundidad semiespacial de un punto $X \in \mathbb{R}^d$ respecto de una muestra X_1, X_2, \dots, X_n en \mathbb{R}^d como la menor fracción de puntos de la muestra que hay en un semiespacio cerrado que contenga a x , es decir,

$$HD_n(x) = \frac{1}{n} \min_{u \in \mathbb{R}^d} \#\{i : \langle X_i, u \rangle \geq \langle x, u \rangle\},$$

donde $\langle \cdot \rangle$ denota el producto escalar. A partir de esta definición, Rousseeuw y Ruts (1996) adaptaron esta noción de profundidad a una distribución de probabilidad F , como el ínfimo de las probabilidades de los semiespacios cerrados que contienen al punto.

$$HD(x, F) = \inf\{P(H) : x \in H \text{ semiespacio cerrado}\}.$$

En el caso univariante y para distribuciones absolutamente continuas, si consideramos P una medida de probabilidad en \mathbb{R} y X una variable aleatoria con función de probabilidad acumulada F , entonces:

$$HD(F, x) = \min\{P(X \leq x), (1 - P(X \leq x))\} = \min\{F(x), 1 - F(x)\}$$

La principal desventaja que presenta esta profundidad es que muy costosa computacionalmente en dimensiones grandes.

TEOREMA 2.1. *La profundidad semiespacial cumple las propiedades $D_1 - D_4$*

Profundidad Simplicial

Liu [1990] define la profundidad simplicial de un punto $x \in \mathbb{R}^d$ respecto a F como la probabilidad de que x se encuentre en el simplex abierto en \mathbb{R}^d , es decir,

$$SD(x, F) = P(x \in S[X_1, \dots, X_{d+1}]), \quad x \in \mathbb{R}^d$$

donde $S[X_1, \dots, X_{d+1}]$ es el simplex abierto cuyos vértices X_1, X_2, \dots, X_{d+1} son observaciones independientes de F .

En el caso univariante y para funciones absolutamente continuas, la profundidad simplicial se calcula como

$$SD(x, F) = 2F(x)[1 - F(x)]$$

Cuando F es desconocida y solo la muestra X_1, X_2, \dots, X_m está dada, la profundidad simplicial muestral se define como:

$$SD_{F_m}(x) = \binom{m}{d+1}^{-1} \sum_{(*)} I(y \in S[X_1, \dots, X_{d+1}])$$

que mide que tan profundo se encuentra x dentro de la nube de puntos $\{X_1, \dots, X_m\}$. $I(\cdot)$ es la función indicatriz; esto es $I(A) = 1$ si A ocurre y $I(A) = 0$ si no. La función $X_m(\cdot)$ denota la distribución empírica de $\{X_1, \dots, X_m\}$ y $(*)$ recorre por todos los posibles subconjuntos de $\{X_1, \dots, X_m\}$ de tamaño $(d+1)$.

Para determinar si un punto $x \in \mathbb{R}^d$ pertenece o no al simplex $S[X_1, \dots, X_{d+1}]$, basta resolver el siguiente sistema de ecuaciones lineales

$$x = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_{d+1} x_{d+1}; \quad \alpha_1 + \alpha_2 + \dots + \alpha_{d+1} = 1 \quad (2.2)$$

Para un simplex no degenerado, el sistema de $d + 1$ ecuaciones con $d + 1$ incógnitas $\alpha_1 + \alpha_2 + \dots + \alpha_{d+1}$ tiene solución única, y x está dentro del simplex sí y solo si $\alpha_1 + \alpha_2 + \dots + \alpha_{d+1}$ son positivos.

TEOREMA 2.2. *La profundidad simplicial cumple las propiedades $D_1 - D_4$*

Demostración. Ver Liu [1990]. □

Profundidad de Mahalanobis

La profundidad de Mahalanobis fue propuesta por Liu and Tang [1996] como una transformación de la distancia de Mahalanobis de cada punto al vector de medias. Sea X un vector aleatorio con función de distribución F en \mathbb{R}^d . Entonces la medida de profundidad de Mahalanobis de un punto x en \mathbb{R}^d con respecto a la distribución F se define como:

$$MhD(x, F) = [1 + (x - \mu_F)^T \Sigma_F^{-1} (x - \mu_F)]^{-1}$$

donde μ_F y Σ_F son el vector de medias y la matriz de varianzas y covarianzas de F respectivamente. La versión muestral se construye a partir de las estimaciones del vector de medias y de la matriz de covarianzas como

$$Mhd(x, F_m) = [1 + (x - \bar{X})^T S^{-1} (x - \bar{X})]^{-1}$$

TEOREMA 2.3. *Sea F simétrica. Entonces la función de profundidad de Mahalanobis cumple las propiedades $D_1 - D_4$ si μ y Σ tienen equivalencia afín y μ_F coincide con el punto de simetría de F .*

Demostración. Ver Zuo and Serfling [2000]. □

La profundidad de Mahalanobis presenta inconvenientes al tener la necesidad de existencia de los dos primeros momentos de la distribución y del hecho de que sus curvas de nivel son elípticas lo cual hace que para distribuciones que no son angularmente simétricas se obtenga valores de profundidad poco representativos.

Profundidad de las proyecciones

Propuesta por Zuo and Serfling [2000], se construye a partir de la mayor discrepancia entre una proyección unidimensional del punto x y la mediana de la misma proyección aplicada al vector aleatorio X . En concreto, se define como: [Cascos et al., 2011]

$$PD(x, F) = \left(1 + \sup_{\|x\|=1} \frac{|\langle u, x \rangle - \text{Me}(\langle u, X \rangle)|}{\text{MEDA}(\langle u, X \rangle)} \right)$$

donde el vector X sigue una distribución F , Me representa la mediana y MEDA es la mediana de las desviaciones respecto a la mediana.

TEOREMA 2.4. *La profundidad de las proyecciones cumple las propiedades $D_1 - D_4$*

Demostración. Ver Zuo and Serfling [2000] □

Si la distribución de probabilidad es angularmente simétrica, la profundidad tiene su máximo en el centro de esta. El punto de mayor profundidad constituye una mediana con un elevado punto de ruptura.

2.2. Técnicas de remuestreo

En muchos procedimientos estadísticos es necesario conocer determinadas características de la distribución muestral de los estadísticos o estimadores empleados. Un enfoque clásico para obtener medidas de exactitud de un estimador es calcularlas mediante análogos empíricos de las fórmulas explícitas.

Sin embargo, en la mayoría de estadísticos es muy difícil (o imposible) obtener fórmulas exactas y explícitas de las medidas de exactitud. Una solución es considerar una aproximación o una expansión asintótica de la medida de exactitud, y estimarla mediante el análogo muestral de la aproximación obtenida, no obstante, este acercamiento presenta claras desventajas. [Alonso et al., 2002]

Los métodos de remuestreo reemplazan las derivaciones teóricas del enfoque anterior por la evaluación de los estadísticos en remuestras o submuestras obtenidas a partir de datos originales, y mediante estos valores se obtienen estimadores de la medida de exactitud o de la distribución muestral del estadístico. Los métodos de remuestreo más populares son el jackknife (Quenouille, 1949) y bootstrap [Efron and Tibshirani, 1986].

2.2.1. Método Bootstrap

El método bootstrap propuesto por Efron and Tibshirani [1986] es una herramienta poderosa para estimar la distribución muestral de un estadístico dado. La idea básica detrás del bootstrap es que la inferencia sobre una población a partir de datos de muestra (muestra \rightarrow población) se pueda modelar volviendo a muestrear los datos de muestra y realizando inferencias sobre una muestra a partir de datos remuestreados (remuestreo \rightarrow muestra). Como la población es desconocida, se desconoce el verdadero error en un estadístico de muestra respecto a su valor poblacional. En remuestras bootstrap, la "población" es de hecho la muestra, y esto es conocido; por lo tanto, la calidad de la inferencia de la muestra "verdadera" de los datos remuestreados (remuestreo \rightarrow muestra) es medible.

La estimación bootstrap de la distribución muestral es generalmente mejor que la aproximación normal basada en el teorema del límite central, aunque el estadístico no esté estandarizado. Sea $\vec{X} = \{X_1, \dots, X_N\}$ una muestra que sigue una distribución F , estamos interesados en hacer inferencia sobre $\theta = \theta(F)$, para ellos debemos conocer la distribución en el muestreo de $R(\vec{X}, F)$, cierto estadístico función de la muestra y de la distribución poblacional (por ejemplo $R = R(\vec{X}, F) = \theta(F_n) - \theta(F)$). El procedimiento bootstrap estándar consiste en reemplazar la distribución poblacional (desconocida) F por una estimación, \hat{F} de la misma (por ejemplo $\hat{F} = F_n$ donde F_n es la distribución empírica) obteniendo condicionalmente de la muestra observada reemplazos de tamaño N de $\{X_1, \dots, X_N\}$. Denotemos la muestra bootstrap como $\{X_1^*, \dots, X_N^*\}$ con distribución \hat{F} . La distribución muestral de $R^* = R(\vec{X}^*, \hat{F})$ bajo \hat{F} es la aproximación bootstrap de la distribución muestral de R bajo F . En raras ocasiones la distribución bootstrap de R^* es calculable directamente, pero siempre puede aproximarse por Monte Carlo.

Ejemplo: Inferencia sobre la media con varianza conocida

Contexto clásico:

$\vec{X} = \{X_1, \dots, X_N\} \sim F$ con μ desconocida y σ conocida.

$$\theta(F) = \mu = \int x dF(x) = \int x f(x) dx$$

$$\theta(F_n) = \int x dF_n(x) = \sum \frac{1}{n} X_i = \frac{1}{n} \sum X_i = \bar{X}$$

$$R = R(\vec{X}, F) = \sqrt{n} \frac{\bar{X} - \mu}{\sigma}$$

Bajo normalidad ($X \sim N(\mu, \sigma^2)$), $R \sim N(0, 1)$. Si F no es normal, tan solo bajo ciertas condiciones, $R \xrightarrow{d} N(0, 1)$.

Contexto bootstrap: $\hat{F} = F_n$

Para cada $i = 1 \dots n$, $P^*(X_i = X_j) = \frac{1}{n}$, $j = 1 \dots n$.
 $\vec{X}^* = (X_1^*, \dots, X_n^*)$
 $R^* = R(\vec{X}^*, \hat{F}) = \sqrt{n} \frac{\bar{X}^* - \bar{X}}{\sigma}$, con $\bar{X}^* = \frac{1}{n} \sum X_i^*$.

Distribución bootstrap aproximada por Monte Carlo

Al conocerse en el bootstrap, el mecanismo que genera los datos, siempre se puede simular dicho mecanismo mediante el método de Monte Carlo ¹. Así un algoritmo general para la aproximación de Monte Carlo del bootstrap sería:

1. Para cada $i = 1, \dots, n$ obtener X_i^* a partir de \hat{F} .
2. Obtener $\vec{X}^* = (X_1^*, \dots, X_n^*)$.
3. Calcular $R^* = R(\vec{X}^*, \hat{F})$.
4. Repetir B veces los pasos 1-3 para obtener las réplicas bootstrap $R^{*(1)}, \dots, R^{*(B)}$
5. Utilizar estas réplicas bootstrap para aproximar la distribución en muestreo de R .

Normalmente el valor de B se toma del orden de varias centenas. En los casos en los que el bootstrap se utiliza para estimar el sesgo o la varianza de un estimador, bastará tomar un número de réplicas bootstrap de tamaño $B = 100, 200, 500$. Sin embargo, cuando se trata de utilizar el bootstrap para contrastes de hipótesis son necesarios valores mayores, por ejemplo $B = 500, 1000, 2000$. [Efron and Tibshirani, 1986]

2.2.2. Bootstrap Suavizado

En el Bootstrap no paramétrico asumimos que la distribución empírica asigna igual masa a cada observación, el bootstrap uniforme es aquel que reemplaza la distribución poblacional desconocida F por la distribución empírica F_n :

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

¹Amplia clase de algoritmos computacionales que se basan en el muestreo aleatorio repetido para obtener resultados numéricos. Su idea esencial es usar la aleatoriedad para resolver problemas que podrían ser deterministas en principio.

Es decir, $\hat{F} = F_n$ y así, $R^* = R(\vec{X}^*, F_n)$. La distribución bootstrap uniforme de R^* es aproximable por Monte Carlo como:

1. Para cada $i = 1, \dots, n$ obtener X_i^* a partir de F_n .
2. Obtener $\vec{X}^* = (X_1^*, \dots, X_n^*)$.
3. Calcular $R^* = R(\vec{X}^*, F_n)$.
4. Repetir B veces los pasos 1-3 para obtener las réplicas bootstrap $R^{*(1)}, \dots, R^{*(B)}$
5. Utilizar estas réplicas bootstrap para aproximar la distribución en muestreo de R .

Para llevar a cabo el paso 1 es necesario simular valores de una variable aleatoria discreta, que toma los valores X_1, \dots, X_n con probabilidad $\frac{1}{n}$. Al tener el caso de equiprobabilidad podemos aplicar el método de la transformación cuantil con búsqueda directa que consiste en simular un número aleatorio U , con distribución $U(0, 1)$, y hacer $X^* = X_{\lceil nU \rceil + 1}$ donde $\lceil x \rceil$ representa la función techo de x . Luego, de acuerdo a este procedimiento podemos escribir al paso 1 como:

1. Para cada $i = 1 \dots, n$ calcular $U_i \sim U(0, 1)$ y hacer $X^* = X_{\lceil nU_i \rceil + 1}$

Al ser F_n discreto, una estimación "suave" de F podría resultar mejor cuando se espera que F sea continuo. La suavización consiste en reemplazar cada punto de los datos con una distribución continua de masa total $\frac{1}{n}$ centrada en el punto. La distribución de suavización más común es uniforme en el intervalo $[-h, h]$. La función de distribución empírica suavizada uniformemente F_n^U es similar a F_n excepto que los saltos de tamaño $\frac{1}{n}$ son reemplazadas por líneas rectas con pendiente $\frac{1}{2nh}$ que pasan por el punto medio del salto, denominado el estimador de Parzen-Rosenblatt definido como:

$$\frac{1}{2nh} = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right) = \frac{1}{n} \sum_{i=1}^n K_h(x - X_i)$$

donde $K_h(u) = \frac{1}{h}K(\frac{u}{h})$, K es una función núcleo (distribución simétrica en torno a cero) y $h > 0$ un parámetro de suavizado llamado ventana. Ahora, eligiendo como función K la densidad de una $U(-1, 1)$, el estimador de Parzen-Rosenblatt resulta:

$$\hat{f}_h = \frac{1}{nh} \sum_{i=1}^n \frac{1}{2} \mathbb{1}\left\{\frac{x - X_i}{h} \in (-1, 1)\right\} = \frac{1}{2nh} \sum_{i=1}^n \mathbb{1}\{X_i \in (x - h, x + h)\}$$

$$= \frac{\#\{X_i \in (x - h, x + h)\}}{2nh}$$

que no es más que la frecuencia relativa de datos X_i en el intervalo $(x - h, x + h)$ dividida entre la longitud del intervalo en cuestión ($2h$).

De tal forma el método bootstrap suavizado se construye como:

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 muestras bootstrap $\vec{X}^* = (X_1^*, \dots, X_n^*)$ a partir de la densidad \hat{f}_h
3. Calcular $R^* = R(\vec{X}^*, \hat{F}_h)$
4. Repetir B veces los pasos 2-3 para obtener las réplicas bootstrap $R^{*(1)}, \dots, R^{*(B)}$

El paso 2 puede relacionarse con la simulación de un valor, V , con densidad K , tal que $X_i + hV$. Así el paso 2 puede escribirse como:

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

En la práctica es razonable elegir h como un valor "bastante pequeño" en relación con la desviación típica de la muestra. Nótese que cuando $h = 0$ este método se reduce al bootstrap uniforme. [Cao, 2017]

2.2.3. Bootstrap estacionario

Métodos de remuestreo independientes idénticamente distribuidos solo funcionan para datos independientes idénticamente distribuidos, métodos de Bootstrap por bloques no paramétricos basados en suposiciones débiles postula que "bloques" pueden ser muestreados de tal manera que pueden ser aproximadamente independientes idénticamente distribuidos.

Bootstrap por bloques

El bootstrap por bloques fue propuesto por Liu and Tang [1996], el cual consiste en remuestrear bloques de observaciones consecutivas en el tiempo. Como resultado, la estructura de dependencia de las observaciones originales se preserva dentro de cada bloque.

Supongamos que $\{X_t\}_{t \in \mathbb{N}}$ es una serie de tiempo estacionaria debilmente dependiente y que $\{X_1, \dots, X_n\}$ son observados. Sea ℓ un entero tal que satisface $1 \leq \ell \leq b$. Definimos los bloques superpuestos B_i como:

$$B_i = \{X_i, \dots, X_{i+\ell-1}\}$$

con $1 \leq i \leq b$ donde $b = n - \ell + 1$. La muestra MBB es obtenida al elegir aleatoriamente un número adecuado de bloques de la colección $\{B_1, \dots, B_b\}$. En consecuencia, denotemos a B_1^*, \dots, B_k^* como la muestra bootstrap extraida con reemplazo de $\{B_1, \dots, B_b\}$. Designando los elementos en B_i^* por $\{X_{i-\ell+1}^*, \dots, X_{i\ell}^*\}$, $i = 1, \dots, k$, la muestra MBB de tamaño $m = k \times \ell$ es X_1^*, \dots, X_m^* . El caso especial en el cada bloque contiene un solo elemento, es decir, $\ell = 1$ es el método bootstrap clásico propuesto por Efron.

De tal forma que el método MBB se construye como:

1. Fijar un entero positivo, ℓ , el tamaño del bloque, y tomar k igual al mayor entero mayor o igual que $\frac{n}{\ell}$.
2. Definir los bloques: $B_i = (X_i, X_{i+1}, \dots, X_{i+\ell-1})$, para $1 \leq i \leq b$ donde $b = n - \ell + 1$.
3. Arrojar k observaciones (bloques), $\zeta_1, \zeta_2, \dots, \zeta_k$ con distribución equiprobable sobre el conjunto de posibles bloques: $\{B_1, B_2, \dots, B_b\}$. Cada ζ_i es un vector ℓ -dimensional $(\zeta_{i1}, \dots, \zeta_{i\ell})$.
4. Definir \vec{X}^* como el vector formado por las n primeras componentes de:

$$(\zeta_{11}, \zeta_{12}, \dots, \zeta_{1\ell}, \dots, \zeta_{k1}, \dots, \zeta_{k\ell})$$

[Carlstein et al., 1998] propusieron una modificación al MBB la cual consiste en seleccionar las remuestras de bloques de acuerdo a una cadena de Markov. El primer bloque de la remuestra bootstrap se genera igual que para el MBB ordinario. Una vez que se selecciona en la remuestra del bloque B_i , el siguiente bloque de la remuestra bootstrap se elige dentro de todos los posibles bloques, B_j , poniendo más probabilidad a aquellos que son precedidos por un bloque, B_{j-1} , cuyo último valor, X_{j+b-2} , es más cercano al último valor, X_{i+b-1} , del bloque B_i . En el caso $j = 1$, esta regla carece de sentido ya que no existe un bloque anterior a B_1 , así que, en este caso la probabilidad depende de la distancia entre X_1 (el primer valor del bloque B_1) y el valor siguiente al último del bloque B_i , es decir X_{i+b} .

Un factor determinante en el bootstrap por bloque es la elección del tamaño de bloque b . Hall et al. [1995] consideran este problema en el contexto de la estimación bootstrap del sesgo y la varianza. Obtienen una expresión asintótica para el error cuadrático medio: $n^{-2}(C_1b^{-2} + C_2n^{-1}b)$, donde C_1 y C_2 son constantes desconocidas que dependen del problema de estimación del que se trate. Está claro entonces que el tamaño óptimo del bloque (en el sentido del error cuadrático medio) es de orden $n^{1/3}$.

Bootstrap estacionario

El bootstrap estacionario difiere del método previo por bloques debido a que usa bloques de longitud aleatoria donde cada longitud de bloque es aproximada por una distribución geométrica en lugar de longitud fija ℓ .

Politis and Romano [1994] propusieron el proceso de remuestreo de datos iniciales a fin de formar nuevas pseudo muestras de series de tiempo y reestimar estadísticos de interés, pero con la diferencia concerniente a estacionariedad. Este procedimiento es generalmente aceptado en la aplicación de datos débilmente dependientes. Entonces, de acuerdo con las observaciones originales $\{X_1, \dots, X_n\}$, las pseudo series temporales $\{X_1^*, \dots, X_n^*\}$ son generadas usando un esquema adecuado creando muestras estacionales. El procedimiento trata de replicar características de la muestra original manteniendo una característica de la estacionariedad dentro de las pseudo muestras de series temporales.

Comenzando por el supuesto de que la muestra original $\{X_1, \dots, X_n\}$ es estrictamente estacionaria y débilmente dependiente. La distribución del estadístico de interés $T_n(X) = T_n\{X_1, \dots, X_n\}$ es estimada. Sea $B_{i\ell} = (X_i, \dots, X_{i+\ell-1})$ el bloque que contiene ℓ observaciones repetidas a partir de X_i . Para garantizar que todas las observaciones iniciales tengan la misma probabilidad de ser extraídas, se sugiere el esquema de bloque circular. Independiente de $\{X_1, \dots, X_n\}$, sean L_1, L_2, \dots la serie de variables aleatorias independientes idénticamente distribuidas con distribución geométrica:

$$P(L_i = m) = (1 - p)^{m-1}p$$

Donde $p \in (0, 1)$ y $p \rightarrow 0, np \rightarrow \infty$. Independientemente de X_i y L_i , sean I_1, I_2, \dots una sucesión de variables uniformes discretas independientes idénticamente distribuidas del conjunto $\{1, \dots, n\}$. Así, pseudo series temporales $\{X_1^*, \dots, X_n^*\}$ son generadas por la sucesión de bloques de longitud aleatoria $B_{I_1, L_1}, B_{I_2, L_2}, \dots$. Primero L_1 observaciones son determinadas usando el primero bloque B_{I_1, L_1} de la serie

$X_{I_1}, \dots, X_{I_1+L_1-1}$, a continuación L_2 observaciones en el bloque B_{I_2, L_2} de la serie $X_{I_2}, \dots, X_{I_2+L_2-1}$. El proceso continúa hasta que alcanza n observaciones dentro de la pseudo-muestra aunque está claro que el proceso mencionado deja espacio para expandir la pseudo-muestra para un número arbitrario de observaciones.

El número mínimo de bloques necesario, k , en el método SB, de forma que el conjunto de bloques $B_{I_1, L_1}, B_{I_2, L_2}, \dots, B_{I_k, L_k}$ tenga al menos n observaciones, coincide con el menor entero k para el cual $\sum_{i=1}^k L_i \geq n$. [Cao, 2017].

Politis and Romano [1994] demostraron la validez asintótica del bootstrap estacionario:

$$\sup_{x \in \mathbb{R}} |P^* \{ \sqrt{n}(\bar{X}_n^* - \bar{X}_n) \leq x \} - P \{ (\bar{X}_n - \mu) \leq x \} | \rightarrow 0,$$

en probabilidad, siempre que $p \rightarrow 0$ y $np \rightarrow \infty$.

2.3. Gráficos de control basados en la profundidad de datos

La mayoría de los gráficos de control están basados en supuestos paramétricos para modelar el comportamiento del proceso, los cuales constituyen un limitante en su aplicación pues en caso de no verificarse, la validez del método queda afectada al igual que las conclusiones que se derivan del mismo.

Dos supuestos claves en la metodología estándar son la normalidad de F y la independencia entre observaciones, la propuesta de Liu [1995] para el diseño de gráficos de control para la Fase II basado en el estadístico Rank es similar a la aproximación asintótica en términos de potencia de la prueba.

Representando las medidas de calidad originales por sus correspondientes rangos univariantes, se pueden construir gráficos de control a partir de estos rangos siguiendo los mismos principios de los gráficos de control convencionales. La ventaja de este acercamiento es que es completamente no paramétrico, así que los gráficos resultantes son válidos sin suposiciones paramétricas en el modelo.

Más aún, estos gráficos nos permiten detectar simultáneamente el cambio de posición y el crecimiento de escala en un proceso.

2.3.1. Estadísticos derivados de la profundidad de datos

Supongamos que d ($d \geq 1$) características son usadas para determinar la calidad de un producto. El proceso es considerado en control si las medidas siguen una distribución de calidad prescrita. Denotemos a F como la distribución k -dimensional prescrita y sean X_1, X_2, \dots, X_m, m observaciones aleatorias de F . La muestra X_1, X_2, \dots, X_m se denomina muestra de referencia y es considerada como la medida de los productos producidos por un proceso en control. Sean Y_1, Y_2, \dots las nuevas observaciones del proceso de manufactura. Supongamos que los Y_i siguen una distribución G , basados en las observaciones Y_i queremos determinar si la calidad del producto se ha deteriorado o de si el producto está fuera de control.

La prueba de hipótesis está dada por:

$$H_0 : F = G$$

H_a : existe un cambio de localización y/o incremento de escala de F a G .

Supongamos que F y G son distribuciones absolutamente continuas, claramente la profundidad de datos induce un ordenamiento de adentro hacia afuera de los puntos de la muestra si los valores de profundidad son calculados y comparados. De manera más específica, si ordenamos los $D_F(X_i)$ de manera ascendente y usamos $X_{[j]}$ para denotar el punto de la muestra asociado con el j -ésimo valor de profundidad más pequeño, entonces $X_{[1]}, X_{[2]}, \dots, X_{[m]}$ son los estadísticos de orden de los X_i , con $X_{[m]}$ su punto más central. Mientras más pequeño sea el orden (o rank) de un punto, más alejado estará ese punto respecto a la distribución subyacente.

Estadístico Rank

Bajo una función de profundidad de datos, definimos al estadístico *rank* $r_F(x)$ que caracteriza la distancia entre F y G con respecto a la profundidad de datos cuando $X \sim F$ y $Y \sim G$, para algún $x \in \mathbb{R}^d$ como:

$$r_F(x) = P\{D_F(X) \leq D_F(x) | X \sim F\} \quad (2.3)$$

Nótese que cuando $D(F, \cdot)$ es afín invariante entonces $r_F(x)$ también lo es.

$$r_F(x) = r_{F_{Ax+b}}(Ax + b)$$

La invarianza afín asegurará que los gráficos de control sean libres de coordenadas. Por otro lado, si F es desconocida, definimos $r_{F_m}(x)$ como la proporción de los X_j

que satisfacen $D(F_m, X_j) \leq D(F_m, x)$

$$r_{F_m}(x) = \#P\{D_{F_m}(X) \leq D_{F_m}(x), j = 1, \dots, m\} / m \quad (2.4)$$

Donde los valores de $D(F_m, \cdot)$ son valores de profundidad empíricos calculados con respecto a F_m .

PROPOSICIÓN 2.5. *Supongamos que $F = G$ y sea $Y \sim G$. Sea $U[0,1]$ una distribución uniforme con soporte en $[0,1]$. Si $D_G(X)$ tiene distribución absolutamente continua, entonces*

1. $r_F(Y) \sim U[0,1]$, y
2. $r_{F_m}(Y) \xrightarrow{\mathcal{L}} U[0,1]$ cuando $m \rightarrow \infty$ a lo largo de casi todas las sucesiones $\{X_1, \dots, X_m\}$ siempre que $D_{F_m}(\cdot)$ converge a $D_F(\cdot)$ uniformemente cuando $m \rightarrow \infty$.

Estadístico Q

El promedio de la proporción $r_F(x)$ para todos los y generados de la población G , denotados por $Q(F, G)$ se calcula como:

$$Q(F, G) = P\{D_F(X) \leq D_F(Y) | Y \sim G, X \sim F\} = E_G[r_F(X)]$$

El estadístico $Q(F, G)$ se denomina *índice de calidad* y toma valores entre $[0,1]$ [Liu and Singh, 1993].

Nótese que cuando $D(F, \cdot)$ es afín invariante entonces $Q(F, G)$ también lo es.

$$Q(F, G) = Q(F_{Ax+b}, G_{Ay+b})$$

Sea $G_n(\cdot)$ la distribución empírica de la muestra $\{Y_1, \dots, Y_n\}$ podemos escribir al índice de calidad Q como:

$$Q(F, G_n) = \frac{1}{n} \sum_{i=1}^n r_F(Y_i), \quad (2.5)$$

El parámetro Q muestra si existe o no una diferencia en la posición y/o dispersión de al comparar F con G . Bajo hipótesis nula $H_0 : F = G$, $Q = \frac{1}{2}$, y $Q < \frac{1}{2}$ indican que existe un posible cambio en la posición o incremento de escala de F a G . Si $Q > \frac{1}{2}$, entonces G tiene una menor dispersión con quizá la misma posición o con cambio relativamente menor en la posición.

Si $X = \{X_1, \dots, X_m\}$ es una muestra aleatoria de la distribución desconocida F y $Y = \{Y_1, \dots, Y_n\}$ es una muestra aleatoria tomada de la distribución G , entonces la estimación de $Q(F, G)$ es:

$$Q(F_m, G_n) = \frac{1}{n} \sum_{i=1}^n r_{F_m}(Y_i). \quad (2.6)$$

PROPOSICIÓN 2.6. *Supongamos que se cumplen los supuestos de la Proposición 2.5 Entonces*

1. $[Q(F, G_n) - \frac{1}{2}] \xrightarrow{\mathcal{L}} \mathcal{N}(0, 1/(12n))$ cuando $n \rightarrow \infty$
2. $[Q(F_m, G_n) - \frac{1}{2}] \xrightarrow{\mathcal{L}} \mathcal{N}(0, [(1/m) + (1/n)]/(12))$ cuando $\min(m, n) \rightarrow \infty$. Si F es una distribución unidimensional y su densidad está acotada superior e inferiormente en una vecindad de la mediana.

2.3.2. Gráficos de control

Gráfico r

Este gráfico es basado en el estadístico rank definido anteriormente.

Se calcula $\{r_F(Y_1), r_F(Y_2), \dots\}$. (o $\{r_{F_m}(Y_1), r_{F_m}(Y_2), \dots\}$ si solo Y_1, \dots, Y_m son conocidos, pero no F). El gráfico r es el gráfico de $r_F(Y_i)$ (o $r_{F_m}(Y_i)$) contra el tiempo i , con $LC=0.5$ y límite de control α , definido como $\alpha = P(Z > z_\alpha)$ donde z_α indica el valor crítico superior α de la distribución normal estándar y $Z \sim \mathcal{N}(0, 1)$.

El proceso es declarado fuera de control si $r_F(\cdot)$ cae bajo α , nótese que α es la tasa de falsas alarmas, la cuál es generalmente cercana a cero, por tanto el gráfico r tiene solamente $LCI=\alpha$ pero no LCS.

La ecuación 2.4 muestra que $r_{F_m}(Y)$ es un índice de qué tan lejano se encuentra Y respecto a la nube de puntos X_i . Un valor pequeño de $r_{F_m}(Y)$ significa que tan solo una pequeña proporción de los X_i están más lejanos que Y . Así, X está en la periferia y no se ajusta a la mayoría de la parte central del conjunto de datos. Suponiendo que $Y \sim G$, un valor pequeño de $r_{F_m}(\cdot)$ sugiere una posible desviación de G a F . Dado que $r_{F_m}(\cdot)$ se define de acuerdo a la profundidad de datos, la posible desviación puede ser causada por un cambio en el "centro" y/o incremento en la escala. [Liu and Singh, 1993].

Por tanto, para el gráfico r con $LCI=\alpha$ corresponde a prueba al nivel α de la siguiente hipótesis:

$$H_0 : F = G$$

H_a : existe un cambio de localización y/o incremento de escala de F a G .

Podemos observar que la hipótesis alternativa es adecuada para detectar el deterioro en la calidad, dado que presenta una pérdida de precisión.

Bajo la hipótesis nula $F = G$ la Proposición 2.5 implica que el valor esperado de $r_F(Y)$ es 0.5 y el de $r_{F_m}(Y)$ es 0.5 casi seguramente para todas las sucesiones $\{X_1, \dots, X_m\}$ para un m grande. Cuando $r_F(Y)$ (o $r_{F_m}(Y)$) tienen valores mucho más pequeños que 0.5, existe evidencia para inclinarse por la H_a , señalando un posible deterioro de la calidad. Por otro lado, cuando $r_F(Y)$ (o $r_{F_m}(Y)$) es más grande que 0.5, es una indicación de una disminución en la escala con quizás un cambio de ubicación insignificante. Esto se ve como una mejora en la calidad, denominado ganancia en precisión, y por lo tanto el proceso no debe verse como fuera de control. La distribución uniforme de $r_F(Y)$ (o $r_{F_m}(Y)$) implica que el límite de control inferior debe ser α .

Así los límites de control para el gráfico r están definidos como: Liu [1995]

$$LC = 0.5$$

$$LCI = \alpha$$

Gráfico Q

Se grafica los promedios de subconjuntos de los $r_F(Y_i)$ (o $r_{F_m}(Y_i)$). Supongamos que cada subconjunto es de tamaño n , los promedios de $r_F(Y_i)$ y $r_{F_m}(Y_i)$ están dados por $Q(F, G_n^j)$ y $Q(F_m, G_n^j)$. Aquí G_n^j denota la distribución empírica de los Y_i en el j -ésimo subconjunto, $j = 1, 2, \dots$. La carta Q grafica

$$\{Q(F, G_n^1, Q(F, G_n^2, \dots)\}$$

o

$$\{Q(F_m, G_n^1, Q(F_m, G_n^2, \dots)\}$$

si solamente X_1, \dots, X_m están disponibles.

Los valores de la LC y LCI dependen de la elección de n . Si n es grande, de acuerdo a la Proposición 2.6 la LC debe ser 0.5, mientras que el LCI debe ser $(0.5 - z_\alpha(12n)^{-1/2})$ para graficar los $\{Q(F, G_n^j)\}$ y $\{0.5 - z_\alpha\sqrt{\frac{1}{12}[(1/m) + (1/n)]}\}$ para graficar los $\{Q(F_m, G_n^j)\}$. Esta aproximación es razonable para n tan pequeños como 5. Sin embargo, para casos de n más pequeños, se puede usar la distribución exacta de $Q(F, G_n)$ dada por la Proposición 2.7 nótese que para pequeños valores de α el

gráfico Q debe tener $LC=0.5$ y $LCI=(n!\alpha)^{1/n}/n$.

PROPOSICIÓN 2.7. Sea $\{U_1, \dots, U_n\}$ una muestra iid de $U[0, 1]$ y sea $H_n(t)$ la función de distribución de $\sum_{i=1}^n U_i$, esto es, $H_n(t) = P(\sum_{i=1}^n U_i \leq t)$. Entonces para cada $n = 1, 2, \dots$, $H_n(t) = 0$ para $t \leq 0$ y

$$H_n(t) = \frac{1}{n!} \sum_{k=0}^n (-1)^k \binom{n}{k} (t-k)_+^n \quad (2.7)$$

donde

$$(x)_+^n = \begin{cases} 0 & \text{si } x \leq 0 \\ x^n & \text{si } x > 0 \end{cases}$$

Cuando n es pequeño, los resultados asintóticos de la Proposición 2.6 pueden no ser aplicables. Puesto que este caso LCI es el α -ésimo cuartil de la distribución de $Q(F, G_n) = \sum_{i=1}^n r_F(Y_i)$, se necesita la distribución del promedio de variables aleatorias uniformes. De donde se sigue de la fórmula de la distribución de la suma de variables aleatorias uniformes dada por la Proposición 2.7.

La ecuación 2.7 fue derivada por Feller (1971) y muestra que $H_n(\cdot)$ es un polinomio por partes, siendo la parte más relevante:

$$\begin{aligned} H_n(t) &= \frac{1}{n!} t^n, \text{ si } 0 \leq t < 1 \\ &= \frac{1}{n!} (t^n - n(t-1)^n), \text{ si } 1 \leq t < 2 \\ &= \frac{1}{n!} \left(n - n(t-1)^n + \frac{n(n-1)}{2} (t-2)^n \right), \text{ si } 2 \leq t < 3 \end{aligned} \quad (2.8)$$

Para determinar el LCI en el gráfico Q para un valor de n pequeño debemos hallar el valor de ω_α tal que $P(\frac{1}{n} \sum_{i=1}^n U_i \geq \omega_\alpha) \equiv \alpha$. La fórmula 2.8 implica que para $\alpha \leq \frac{1}{n!}$, $n(\omega_\alpha)^n/n! = \alpha$. Consecuentemente $\omega_\alpha = (n!\alpha)^{1/n}/n$. Esto justifica la elección del LCI para el gráfico Q. La solución para esta ecuación es única dado que $H_n(\cdot)$ es una función estrictamente creciente.

2.3.3. Simulación

Se usan datos bivariantes para ilustrar la construcción de los gráficos de control. Sea $F \sim \mathcal{N} \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$. Generamos 540 puntos de muestra de F , denotando los primeros 500 como X_1, \dots, X_{500} y los últimos 40 como Y_1, \dots, Y_{40} . También genera-

mos 40 puntos de muestra de la distribución $\mathcal{N}\left(\begin{pmatrix} 2 \\ 2 \end{pmatrix}, \begin{pmatrix} 4 & 0 \\ 0 & 4 \end{pmatrix}\right)$. y denotamos a estos puntos como Y_{41}, \dots, Y_{80} .

Nótese que existe un cambio de media y crecimiento en la escala en la distribución de los últimos 40 Y_i . Para cada Y_i calculamos su profundidad simplicial. Basados en estos valores calculamos los valores del estadístico rango ($r_{F_m}(Y_i)$).

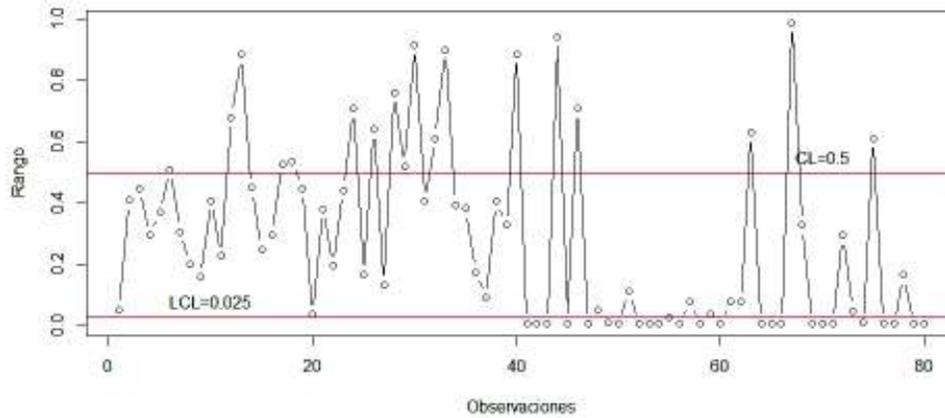


Figura 2.2: Gráfico r

La figura 2.2 muestra el gráfico de los $r_{F_m}(Y_i)$ con $LC=0.5$ y $LCL=0.025$, el cual es el valor escogido de α para todos los gráficos. Podemos ver que el proceso está fuera de control en la segunda mitad debido a que la mayoría de los $r_{F_m}(Y_i)$ caen bajo el LCL. Las falsas alarmas en la primera mitad de los Y_i se atribuyen a fluctuaciones aleatorias de la misma manera que las falsas alarmas están caracterizadas al gráfico X univariante.

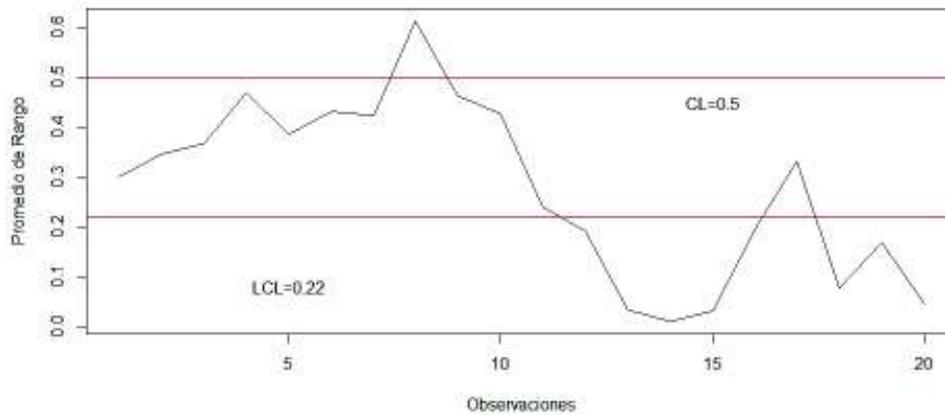


Figura 2.3: Gráfico Q para $n=4$

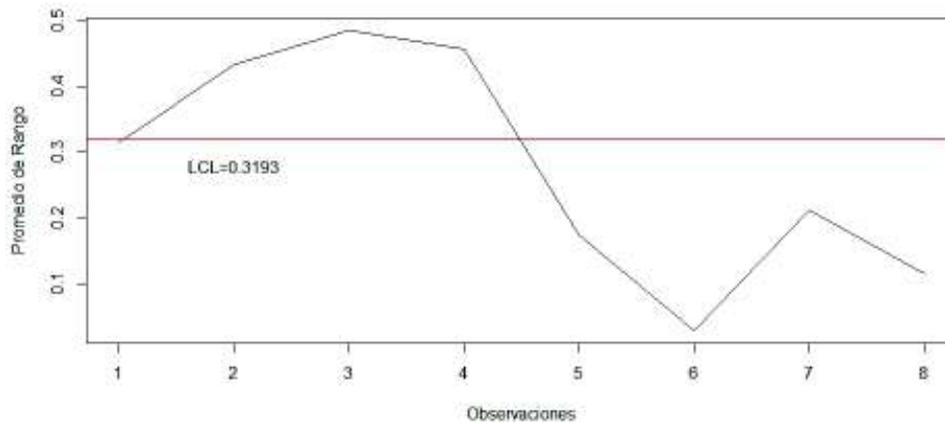


Figura 2.4: Gráfico Q para $n=10$

Las figuras 2.3 y 2.4 muestran el gráfico Q para tamaños de grupo $n = 4$ y $n = 10$ respectivamente. Los $\{Q(F_m, G_n^j), j = 1, 2, \dots\}$ se calculan de acuerdo a la definición. Para la figura 2.3 se elige $LCL=0.22$ de acuerdo a la Proposición 2.7. Para la figura 2.4 se elige $LCL=0.3193$ de acuerdo a la Proposición 2.6. Podemos notar en ambos gráficos que el proceso está fuera de control en la segunda mitad. Podemos observar también que los promedios de los $r_{F_m}(Y_i)$ eliminan las fluctuaciones aleatorias que aparecen en la primera mitad del gráfico r en la figura 2.2.

2.4. Gráficos de control basados en métodos de remuestreo

2.4.1. Gráfico para muestras i.i.d. basadas en el método bootstrap

Liu and Tang [1996] proponen una estimación bootstrap de la distribución muestral la cual sobrepasa la aproximación normal basada en el teorema del límite central a pesar que el estadístico no esté estandarizado. Sea $\{X_1, \dots, X_n\}$ una muestra i.i.d. que sigue una distribución F con media μ y varianza σ^2 . Denotemos la muestra bootstrap como $\{X_1^*, \dots, X_n^*\}$ y la muestra y desviación estándar por \bar{X}_n^* y S_n^* respectivamente. Sea F_n la distribución empírica de $\{X_1, \dots, X_n\}$. La distribución muestral de $(\bar{X}_n^* - \bar{X}_n)$ bajo F_n es la aproximación bootstrap de la distribución muestral de $(\bar{X}_n - \mu)$ bajo F .

PROPOSICIÓN 2.8. *Si el segundo momento de X existe entonces*

$$\left\| P\left(\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \leq x | F_n\right) - P\left(\sqrt{n}(\bar{X}_n - \mu) \leq x | F\right) \right\| \rightarrow 0$$

casi seguramente cuando $n \rightarrow \infty$

Es decir la distribución muestral de $\sqrt{n}(\bar{X}_n - \mu)$ puede aproximarse por su contraparte bootstrap. Supongamos que $n = kN$, con N el tamaño de subgrupo y k el número de subgrupos. Puesto que la carta \bar{X} grafica las medias muestrales de los subgrupos \bar{X}_N , los límites de control se obtienen a partir de los cuartiles $\alpha/2$ y $1 - \alpha/2$ de la distribución muestral de $\sqrt{N}(\bar{X}_N - \mu)$. De hecho, esta distribución muestral puede aproximarse por bootstrap, notamos que para todo x

$$P(\sqrt{N}(\bar{X}_N^* - \bar{X}_N) \leq x | F_n) \approx P(\sqrt{N}(\bar{X}_N - \mu) \leq x | F)$$

Este acercamiento permite construir un gráfico de control \bar{X} de manera distinta; repetir el procedimiento bootstrap K veces y formar un histograma con los K términos resultantes de $\sqrt{N}(\bar{X}_N^* - \bar{X}_N)$, y luego localizar los cuartiles $\alpha/2$ y $1 - \alpha/2$. Estos son usados como una estimación de los cuartiles $\alpha/2$ y $1 - \alpha/2$ de $\sqrt{N}(\bar{X}_N - \mu)$. En otras palabras, usando el histograma bootstrap, obtenemos $\tau_{\alpha/2}$ tal que

$$\alpha/2 = P(\sqrt{N}(\bar{X}_N^* - \bar{X}_N) \leq \tau_{\alpha/2} | F_n), \text{ c.t.p}$$

Esto lleva a

$$\alpha/2 \approx P(\bar{X} \leq \mu + \tau_{\alpha/2}/\sqrt{N} | F)$$

Finalmente obtenemos los límites de control para el gráfico \bar{X}

$$LCI = \bar{\bar{X}} + \tau_{\alpha/2}/\sqrt{n}$$

y

$$LCI = \bar{\bar{X}} + \tau_{1-\alpha/2}/\sqrt{n}$$

usando $\bar{\bar{X}}$ como un estimado de μ bajo H_0 .

Simulación

Comparamos los límites de control obtenidos mediante aproximaciones normales y bootstrap, para esto generamos dos distribuciones: $\mathcal{N}(0,1)$ y una distribución exponencial de parámetro $\lambda = 1$. En cada caso se extrae una muestra aleatoria de tamaño 100 y se forma un subgrupo por cada cuatro observaciones consecutivas, es decir, $N = 4$ y $k = 25$. Aquí $\bar{\bar{X}}$ es el promedio de $\{X_1, \dots, X_{100}\}$, y \bar{X} es el promedio dentro de un subgrupo. De $\{X_1, \dots, X_{100}\}$, obtenemos 200 muestras bootstrap (cada una de tamaño 4) y calculamos los 200 valores correspondientes de $\sqrt{N}(\bar{X} - \bar{\bar{X}})$.

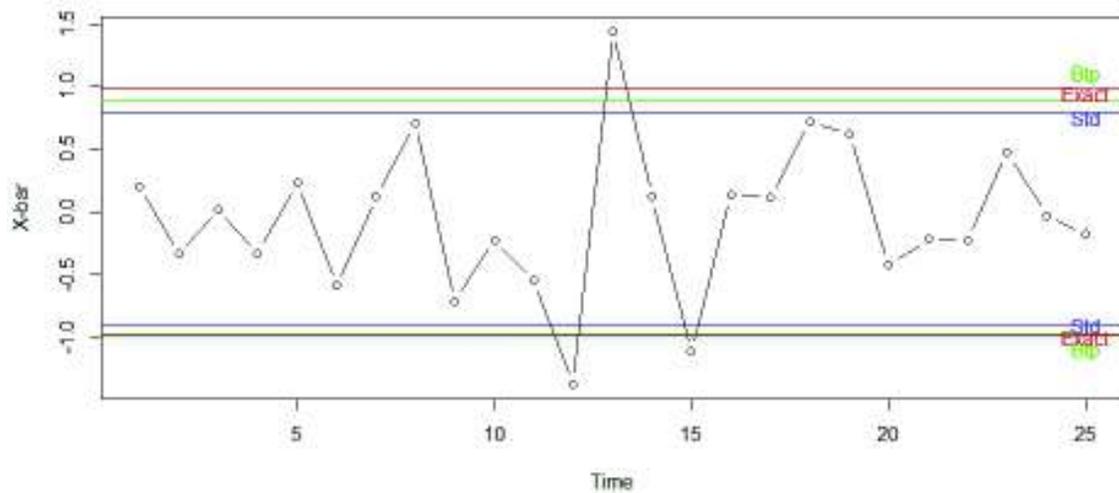


Figura 2.5: Comparación bajo distribución normal ($\alpha = 5\%$)

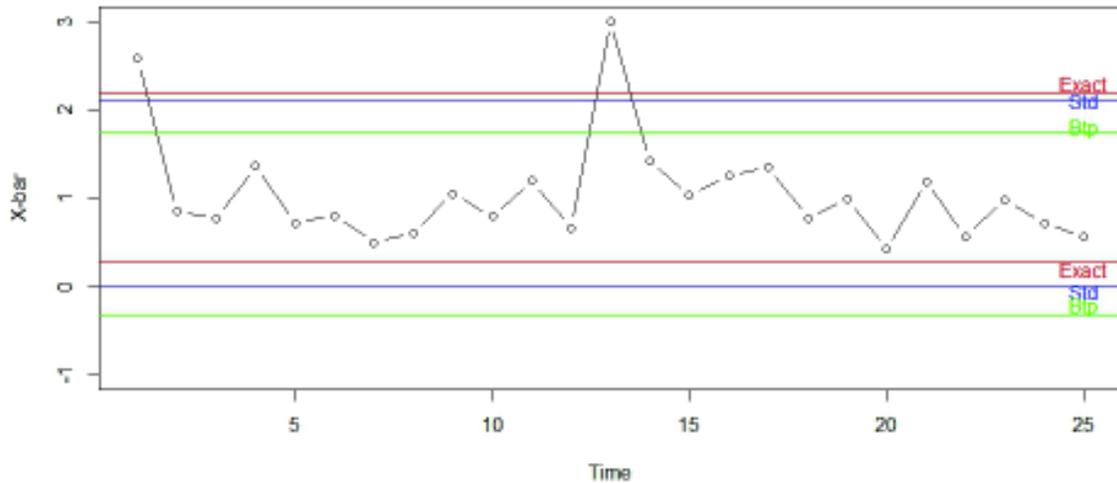


Figura 2.6: Comparación bajo distribución exponencial ($\alpha = 5\%$)

Para el caso de la distribución normal, se esperaría que la aproximación normal estándar supere al método bootstrap, puesto que este método está diseñado específicamente bajo el supuesto de normalidad. Sin embargo, los resultados muestran que el bootstrap compite bien bajo el modelo normal. Para el caso de la distribución exponencial que está sesgada con una cola derecha larga, el gráfico 2.6 indica que el bootstrap es más efectivo en capturar el sesgo de la distribución subyacente. Es decir el método bootstrap puede dar una mejor aproximación de la distribución muestral de un estadístico. En resumen, los límites de control obtenidos por bootstrap son superiores a los obtenidos por aproximaciones normales.

2.4.2. Gráficos de control para observaciones correlacionadas

Un supuesto para la implementación de gráficos de control asume que las observaciones no presentan una estructura de autocorrelación. La suposición de independencia será incumplida de forma cada vez más frecuente en la práctica, en algunas ocasiones, por la presencia de equipos de medición más sofisticados, que permiten obtener un mayor volumen de información en un intervalo de tiempo más pequeño, favoreciendo la aparición de autocorrelación en las observaciones. En otros casos, la autocorrelación va a ser un comportamiento inherente al proceso, que debe ser asumido como un elemento más en su modelización y análisis.

La independencia entre observaciones implica que las observaciones actuales no proporcionan ninguna información del siguiente valor de los datos. Un proceso que

viola el requisito de independencia con frecuencia causaría que la variación a corto plazo sea un indicador deficiente de la variación a largo plazo, generaría límites estrechos y disminuiría la capacidad de detección. Cuando existe correlación entre observaciones, la verdadera varianza del proceso implica covarianzas entre observaciones que no son bien estimadas por la varianza muestral, por tanto, gráficos univariantes ya no son válidos, dado que no es posible cuantificar el riesgo. Métodos de remuestreo independientes idénticamente distribuidos solo funcionan para datos independientes idénticamente distribuidos.

Una sucesión de datos $\{X_1, X_2, \dots\}$ se dice m -dependiente si los datos son dependientes, a menos que sus índices de tiempo estén separados al menos m unidades. Generalmente m -dependencia modela correctamente medidas de calidad, puesto que observaciones muestrales de un proceso de manufactura son generadas en espacios de tiempo igualmente espaciados donde se espera que exista correlación entre observaciones vecinas. Cuando las observaciones son dependientes, las aproximaciones normales no funcionan debido a que se sobrestima (subestima) la varianza de la media del proceso si las observaciones están negativamente (positivamente) correlacionadas. Además de la m -dependencia, suponemos que las observaciones $\{X_1, X_2, \dots\}$ son estacionarias, esto es, que la distribución de un subconjunto de observaciones de X_i no cambia cuando los índices son desplazados el mismo número de unidades. Por ejemplo, (X_3, X_5, X_{10}) tiene la misma distribución de (X_6, X_8, X_{13}) .

Gráficos de control usando MBB

Denotemos a \bar{X}_n^* y \bar{X}_n como la media de la muestra bootstrap y la media de la muestra original respectivamente. Como en el caso iid, debemos obtener la distribución muestral de $\sqrt{N}(\bar{X} - \mu)$ para obtener los límites de control del gráfico \bar{X} . Bajo m -dependencia, $m \leq N$ se puede mostrar que la verdadera varianza de $\sqrt{N}(\bar{X} - \mu)$ es $\sigma_N^2 \equiv \sigma^2 + 2 \sum_{j=1}^m [1 - (j/N)] \text{cov}(X_1, X_{1-j})$ [Liu and Tang, 1996]. Claramente, un buen estimado de σ_N^2 nos permitirá obtener límites de control apropiados. Las dos proposiciones siguientes, dadas por Liu and Singh [1993], muestran como el MBB provee un buen estimado de σ_n^2 y como la distribución muestral de $\sqrt{N}(\bar{X} - \mu)$ puede ser aproximado por su contraparte bootstrap.

PROPOSICIÓN 2.9. *Sea $\{X_1, X_2, \dots\}$ una sucesión de observaciones estacionarias m -dependientes. Si el segundo momento de los X_i es finito, entonces*

$$\left\| P \left(\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \leq x | F_n \right) - \Phi(x/\sigma_\ell) \right\| \rightarrow 0$$

casi seguro cuando $N \rightarrow \infty$ para un tamaño de bloque fijo ℓ . $\Phi(\cdot)$ denota la distribución normal estándar acumulada.

PROPOSICIÓN 2.10. Sea $\{X_1, X_2, \dots\}$ una sucesión de observaciones estacionarias m -dependientes. Si el $(4 + \delta)$ -ésimo momento de los X_i es finito con $\delta > 0$, el tamaño de bloque $\ell \rightarrow \infty$ y $\ell/\sqrt{n} \rightarrow 0$ cuando $n \rightarrow \infty$, entonces

$$\left\| P\left(\sqrt{n}(\bar{X}_n^* - \bar{X}_n) \leq x | F_n\right) - P\left(\sqrt{n}(\bar{X}_n - \mu) \leq x | F\right) \right\| \rightarrow 0$$

en probabilidad cuando $n \rightarrow \infty$.

Si tomamos $\ell = N$ en la Proposición 2.9, se puede ver que la distribución muestral bootstrap converge a una distribución normal con varianza σ_N^2 , la verdadera varianza de $\sqrt{N}(\bar{X} - \mu)$, la cual es necesaria para obtener límites de control correctos. La distribución muestral se puede aproximar mediante bootstrap tal que $\forall x$:

$$P\left(\sqrt{N}(\bar{X}_N^* - \bar{X}_n) \leq x | F_n\right) \approx P\left(\sqrt{N}(\bar{X}_N - \mu) \leq x | F\right)$$

Para obtener los límites de control, repetimos el procedimiento MBB con $\ell = N$, K veces para obtener K términos de \bar{X}_N^* . Localizamos los $(\alpha/2)$ -ésimo y $(1 - \alpha/2)$ -ésimo cuartiles del histograma formado por los K valores de $\sqrt{N}(\bar{X}_N^* - \bar{X})$, y los denotamos por $\tau_{\alpha/2}$ y $\tau_{1-\alpha/2}$. Luego, usamos estos valores como los estimados de los $(\alpha/2)$ -ésimo y $(1 - \alpha/2)$ -ésimo cuartiles de la distribución $\sqrt{N}(\bar{X} - \mu)$ tal que

$$P(\tau_{\alpha/2} \leq \sqrt{N}(\bar{X} - \mu) \leq \tau_{1-\alpha/2}) \approx 1 - \alpha$$

Finalmente, esto sugiere que eligamos

$$\text{LCI} = \bar{X}_n + \tau_{\alpha/2}/\sqrt{N}$$

y

$$\text{LCS} = \bar{X}_n + \tau_{1-\alpha/2}/\sqrt{N}$$

Simulación

Liu and Tang [1996] propusieron simular una muestra aleatoria de 101 puntos de una distribución $\mathcal{N}(0, 1)$ y lo denotamos por $\{A_1, \dots, A_{101}\}$. Luego, obtenemos independientemente una muestra de 100 observaciones de una distribución $\mathcal{N}(0, 0.1)$ y lo denotamos por $\{y_1, \dots, y_{100}\}$. Sea $X_i = (A_i + A_{i+1})/2 + y_i$, $i = 1, \dots, 100$, esto es $n = 100$. Claramente los X_i son 1-dependientes, y X_i sigue una distribución

$\mathcal{N}(0, 0.51)$ con $cov(X_i, X_{i+1}) = 0.25$. Repetimos el procedimiento MBB con $\ell = 4$ ($= N$) 1000 veces para obtener 1000 promedios de muestras bootstrap \bar{X}_N^* y construimos un histograma de los 1000 valores de $\sqrt{N}(\bar{X}_N^* - \bar{X}_n)$. Aquí \bar{X}_n es el promedio de los 100 primeros X_i .

Si suponemos que $\alpha = 0.05$, debemos localizar los 2.5-ésimos y 97.5 cuartiles del histograma. Finalmente, se calcula los límites de control para $\ell = 4$, $\ell = 2$ y $\ell = 1$, además de los límites de control de Shewhart estándar basados en S .

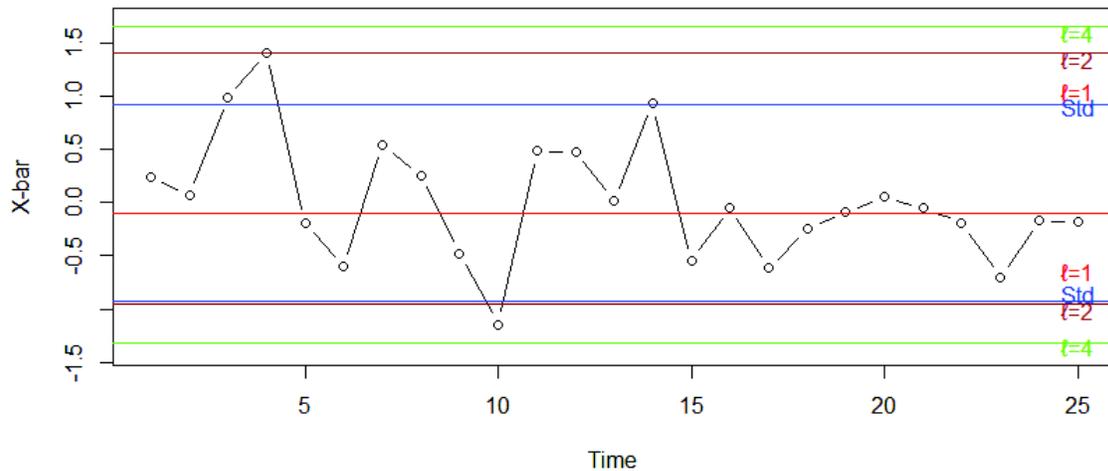


Figura 2.7: Gráfico X-barra para datos dependientes

En el gráfico 2.7 podemos notar que los límites de Shewhart (denotados Std) son más angostos que los calculados mediante MBB, esto es debido a que S^2 calcula la covarianza entre X_i y X_{i+1} en la dirección contraria, por tanto subestima la verdadera varianza.

Podemos agregar también que mientras ℓ decrece en valor, los límites de control correspondientes se mueven hacia adentro lo que provoca que el proceso hubiese estado fuera de control más a menudo de lo que debería, esto es debido a la correlación positiva. Si los X_i estuvieran correlacionados negativamente, los límites de control se moverían hacia afuera, dejando las observaciones fuera de control sin detectar.

2.4.3. Gráficos de control usando SB

Scepi and Acconcia [1996] proponen un marco de referencia no paramétrico que

reemplaza la región de control elíptica multivariante con una empírica convexa, la cual se deriva al combinar el bootstrap estacionario con un método de tres vías para análisis de datos.

Se genera mediante bootstrap estacionario B matrices de datos denotadas por \mathbf{T}_b las cuales pueden ser consideradas como “rodajas de una matriz de tres vías y así analizadas por un acercamiento *interestructura-compromiso-intraestructura* (STRATIS; Escoufier, 1987) con el objetivo de construir gráficos de control no paramétricos multivariantes para un proceso estadístico secuencial.

Sea \mathbf{T}_b ($b = 1, \dots, B$) una matriz de datos que consiste de Q variables medidas en T observaciones. De acuerdo a STATIS, asociamos a cada \mathbf{X}_b el elemento característico \mathbf{O}_b definido como la matriz de productos escalares: $\mathbf{O}_b = \mathbf{T}_b \mathbf{T}_b^T$. El procedimiento de control propuesto se basa en el síntesis por medias de dos matrices de los \mathbf{O}_b . La primera matriz denominada matriz interestructura \mathbf{IS} , con elemento general, is_{bbT} , la distancia entre \mathbf{O}_b y \mathbf{O}_b^T

$$is_{bbT} = \frac{\text{Tr}(\mathbf{O}_b \mathbf{O}_b^T)}{\sqrt{\text{Tr}(\mathbf{O}_b)^2 \text{Tr}(\mathbf{O}_b^T)^2}}$$

La segunda matriz, denominada matriz de compromiso \mathbf{CO} es la matriz más relacionada a cada \mathbf{O}_b . Se calcula como la suma ponderada de los \mathbf{O}_b con el b -ésimo elemento del vector propio, u , correspondiendo al mayor valor propio de \mathbf{IS} :

$$\mathbf{CO} = \sum_{b=1}^B u_b \mathbf{O}_b$$

Al diagonalizar \mathbf{IS} , podemos obtener una representación gráfica de los puntos réplica que representan las muestras de control. En particular, el primer plano principal es posible derivar el primer gráfico de control multivariante. *El gráfico de control IS* es la región de confianza empírica es la región de confianza empírica construida mediante (Greenacre, 1984) convex hull peeling de la nube de los B puntos. Cuando el primer plano principal no satisface en detectar un punto fuera de control, se pueden construir otros gráficos de control \mathbf{IS} en los planos factoriales posteriores para investigar la variabilidad restante.

Al diagonalizar \mathbf{CO} , se obtiene un plano para construir el segundo gráfico de control multivariante: el *El gráfico de control CO*. En el primer plano principal asociado a la matriz \mathbf{CO} , podemos proyectar los elementos T de cada matriz para derivar una región de confianza empírica para cada observación.

El objetivo de estos gráficos de control es distinto. El gráfico de control **IS** permite tener una evaluación general del comportamiento de una muestra $X^+(T, Q)$. Si la proyección del elemento característico O^+ como suplementario en este gráfico, está fuera de la región, tenemos una señal fuera de control para la muestra. Por otro lado, el gráfico de control **CO** permite controlar una sola observación. Si la proyección de una nueva observación t_i está fuera de la región correspondiente, obtenemos una señal fuera de control para esa observación. Los gráficos de control pueden utilizarse conjuntamente para detectar, mediante el gráfico de control **CO**, qué unidades son responsables de un fuera de control registrado por el gráfico de control **IS**.

Capítulo 3

Propuesta de gráficos de control para la Fase I y II

Detectar y evaluar de forma más rápida cualquier anomalía y monitorear un proceso industrial para controlar su buen funcionamiento son algunos de los usos de los gráficos de control. Sin embargo, el aumento de la automatización provoca que la información acerca del estado de los procesos se obtenga a tasas mayores, de tal forma que el desarrollo de nuevos métodos para evaluar estas características es fundamental debido a que son una herramienta contrastada para mejorar la productividad, ya que, al reducir el rechazo y la reelaboración, los costos disminuyen y la capacidad de producción aumenta.

En este capítulo se propone una metodología para el diseño de gráficos de control no paramétricos para la Fases I y II utilizando técnicas de remuestreo a partir de observaciones independientes y autocorrelacionadas. Este diseño se basa en una adaptación de un proceso para detección de datos atípicos en análisis de datos funcionales (Flores, 2018).

La propuesta del diseño de un gráfico de control se basa en la detección de puntos atípicos mediante el cálculo de profundidad de datos y técnicas de remuestreo Bootstrap para datos independientes y dependientes, y la adaptación de los gráficos no paramétricos de profundidades. En particular, se propone una metodología de control que proporciona herramientas para llevar a cabo la Fase II de un gráfico de control. Es decir, se propone la aplicación de métodos con el objeto de eliminar atípicos basados en la profundidad de datos para la eliminación de las principales causas asignables de la variabilidad del proceso causante de errores y anomalías también denominada Fase I. A continuación, a partir de los valores de profundidad de

datos de la muestra de calibrado resultante, se propone la aplicación del gráfico no paramétrico de profundidades para llevar a cabo la Fase II o de monitoreo, es decir, el contraste de si una nueva observación pertenece a la población estimada en la Fase I.

3.1. Metodología

3.1.1. Gráficos de control para calibrado

En esta sección se propone el diseño de un gráfico de control para la Fase I que realiza un análisis retrospectivo de datos históricos para establecer la estabilidad del proceso. Cabe recalcar que el éxito de la fase de monitoreo se basa en el éxito del análisis de la Fase I.

La aplicación de la Fase I consiste en detectar y eliminar datos que tengan desviación. Se desea estimar el cuantil a un nivel α de significancia de la distribución de las profundidades, dado por el límite de control inferior (LCI) y el porcentaje de rechazo basado en la probabilidad de error Tipo II (β) dado por la potencia de la prueba ($1-\beta$).

Lanhede [2015] describe contrastes estadísticos para la Fase I basados en hallar cambios distribucionales en datos ordenados en el tiempo $\mathbf{X} = \{X_1, X_2, \dots, X_m\}$.

Usualmente hay dos escenarios a considerarse, *variación aislada* y *puntos de cambio*.

Una variación aislada es la ocurrencia de al menos una observación de la variable observada que se desvía de la distribución de las otras observaciones, es decir, que el proceso experimentó un evento temporal. La hipótesis usada para la prueba es:

$$H_0 : X_i \stackrel{d}{=} X_j \quad \forall i, j \in \{1, \dots, m\}$$

$$H_a : X_i \stackrel{d}{\neq} X_j \quad \exists i, j \in \{1, \dots, m\}$$

El punto de cambio hace referencia al evento que causa un cambio más duradero en la distribución del proceso. La hipótesis para el caso de k puntos de cambio se formula como:

$$H_0 : X_i \stackrel{d}{=} X_j \quad \forall i, j \in \{1, \dots, m\}$$

$$H_a : \begin{cases} X_i \neq^d X_j & \forall i, j \in \{1, \dots, \tau_1\} \\ X_i \neq^d X_j & \forall i, j \in \{\tau_1 + 1, \tau_1 + 2, \dots, \tau_2\} \\ \vdots & \vdots \\ X_i \neq^d X_j & \forall i, j \in \{\tau_k + 1, \tau_k + 2, \dots, m\} \end{cases}$$

Hay dos factores de rendimiento a considerar cuando se comparan los métodos de Fase I.

- La probabilidad de señal en control, la cual es la misma que el nivel de significancia de la prueba. Una probabilidad baja de error tipo I es deseada.
- La probabilidad de señal fuera de control, la cual es la potencia para detectar cambios distribucionales debe ser lo más alto posible. Sin embargo, la potencia, o la probabilidad de señal fuera de control, de un método depende tanto de la distribución en control como en la distribución fuera de control. Por lo tanto, es necesario estimar la probabilidad de señal para una variedad de escenarios fuera de control con el fin de comparar diferentes de métodos, fortalezas y debilidades.

Para la Fase I, sea F una distribución absolutamente continua, se consideran observaciones dadas por la variable aleatoria X de la cual se extrae una muestra $\{X_1, X_2, \dots, X_n\}$ de donde se ejecutan los siguientes pasos para el diseño del gráfico de control:

Caso independiente

1. Se obtienen las profundidades $D_F(X_i)$ ordenadas descendientemente y se desarrolla un gráfico en función de las profundidades y el tiempo.
2. Se calcula el límite de control inferior (LCI) como el cuantil al nivel α de las observaciones bajo hipótesis nula tal que el porcentaje de falsas alarmas (valores atípicos), son aproximadamente iguales a α , y el porcentaje de rechazo $(1 - \beta)$ como el porcentaje promedio de las observaciones bajo el límite de control. Se incorporan los siguientes procedimientos para el cálculo del LCI:

- Procedimiento bootstrap
 - Se obtienen B muestras denotadas por X_i^b ($i = 1, \dots, n, b = 1, \dots, B$) mediante bootstrap suavizado con coeficiente h a partir del conjunto de datos generado mediante simulación Monte Carlo.

- Para cada $b = 1, \dots, B$ se obtiene el cuantil $C = LCI$ de la distribución de las profundidades $D_F(X_i^b)$ al nivel α .
 - Procedimiento bootstrap basado en recorte
 - Se supone que máximo un porcentaje de la muestra se puede considerar atípico.
 - Se descarta una muestra de tamaño $\alpha * N$ del conjunto de datos simulado.
 - Del conjunto de datos de tamaño $(1 - \alpha) * N$ se extrae una muestra de tamaño $\alpha * N$ la cual forma un nuevo conjunto de datos X_i .
 - Se obtienen B muestras denotado X_i^b mediante bootstrap suavizado con coeficiente de suavización γ a partir del conjunto de observaciones generado.
 - Para cada $b = 1, \dots, B$ se obtiene el cuantil $C = LCI$ de la distribución de las profundidades $D_F(X_i^b)$ al nivel α .
3. Si $D(X_i) \leq LCI$ entonces el proceso está fuera de control.

Caso dependiente

1. Se obtienen las profundidades $D_F(X_i)$ ordenadas descendientemente y se desarrolla un gráfico en función de las profundidades y el tiempo.
2. Se calcula el límite de control inferior (LCI) como el cuantil al nivel α de las observaciones bajo hipótesis nula tal que el porcentaje de falsas alarmas (valores atípicos), son aproximadamente iguales a α , y el porcentaje de rechazo $(1 - \beta)$ como el porcentaje promedio de las observaciones bajo el límite de control. Se incorporan los siguientes procedimientos para el cálculo del LCI:
 - Procedimiento bootstrap
 - Se obtienen B muestras denotadas por X_i^b ($i = 1, \dots, n, b = 1, \dots, B$) mediante bootstrap estacionario utilizando una longitud de bloque l (media de una distribución geométrica) a partir del conjunto de datos generado mediante simulación Monte Carlo.
 - Para cada $b = 1, \dots, B$ se obtiene el cuantil $C = LCI$ de la distribución de las profundidades $D_F(X_i^b)$ al nivel α .
 - Procedimiento bootstrap basado en recorte

- Se supone que máximo un porcentaje de la muestra se puede considerar atípico.
 - Se descarta una muestra de tamaño $\alpha * N$ del conjunto de datos simulado.
 - Del conjunto de datos de tamaño $(1 - \alpha) * N$ se extrae una muestra de tamaño $\alpha * N$ la cual forma un nuevo conjunto de datos X_i .
- Se obtienen B muestras denotado X_i^b mediante bootstrap estacionario con longitud de bloque l a partir del conjunto de observaciones generado.
- Para cada $b = 1, \dots, B$ se obtiene el cuantil $C = LCI$ de la distribución de las profundidades $D_F(X_i^b)$ al nivel α .

3. Si $D(X_i) \leq LCI$ entonces el proceso está fuera de control.

El cálculo del límite de control inferior se basa en una estimación de la distribución de las profundidades mediante técnicas de remuestreo. Nótese que dado que el proceso debe estar bajo control al momento de concluir la Fase I, cuando se detectan datos atípicos estos se eliminan de las muestras hasta que el proceso se encuentre estable.

3.1.2. Gráficos de control para monitoreo

La Fase II se encarga del monitoreo del proceso; usando los datos de la primera fase, la distribución en control del proceso puede ser estimada. Y esta distribución se usa para construir límites de control para el proceso supervisado.

Los límites abarcan un intervalo que debe cubrir nuevas observaciones con alta probabilidad dado que el proceso está en control. Esto cambia el escenario de pruebas, de detectar puntos de cambio y variaciones aisladas en una secuencia de datos observados, a encontrar desviaciones distribucionales entre datos de Fase II recién recolectados y datos de referencia. Sea F la distribución de donde $\mathbf{X} = \{X_1, X_2, \dots, X_m\}$ (la muestra de referencia) fue observada.

De la Fase I se asume que F es la distribución en control del proceso. Por otro lado, sea G la distribución de \mathbf{Y} , una muestra recogida durante el monitoreo de la Fase II. Este método se basa en contrastar la hipótesis:

$$H_0 : F = G$$

$$H_a : F \neq G$$

Como en la Fase I, el rendimiento en control y fuera de control se deben tomar en cuenta. Estos rendimientos se miden en ARL_0 y ARL_1 , el número promedio de corridas cuando el proceso está en control (ARL_0) y fuera de control (ARL_1). El ARL_1 debe ser lo más bajo posible para incrementar la probabilidad de identificar rápidamente eventos fuera de control.

Para la fase de monitoreo se aplican gráficos de control basados en medidas de profundidad presentadas en el capítulo 2. Estos gráficos se contrastan con los gráficos de rangos propuestos por Liu [Liu, 1995]. Usando los datos de la primera fase, la distribución en control del proceso puede ser estimada. El propósito de la Fase II es detectar alguna desviación del comportamiento estimado. Esto se logra recolectando datos del proceso preferentemente en tiempo real y probando la hipótesis de cambio distribucional en el proceso.

El procedimiento para realizar gráficos de control para la Fase II es muy similar al de la Fase I, con la diferencia que no se debe estimar la distribución del estadístico, es decir, el límite de control inferior, ya que fue calculado en la fase previa.

La metodología es la siguiente:

1. A partir de la muestra de referencia $\{X_1, X_2, \dots, X_n\}$ se calcula la profundidad del conjunto de datos $D(X_i)$ con $i = 1, \dots, n$ y con base a esta se obtiene las profundidades de la muestra de monitoreo $D(Y_j)$ con $j = n + 1, \dots, m$ basadas en la muestra de calibrado.
2. Graficar los valores de las profundidades obtenidas contrastando con el límite de control inferior obtenido en la Fase I.
3. Monitorear el proceso, si se tiene observaciones $D(Y_j) \leq LCI$ entonces el proceso está fuera de control.
4. Calcular el porcentaje de rechazo $1 - \beta$ como el promedio de observaciones bajo el límite de control inferior y comparar contra el propuesto por Liu.

Cabe recalcar que en el estudio de simulación se contrastará el rendimiento de los gráficos propuestos contra el propuesto por Liu en donde se determinará la capacidad de detección basado en la potencia de la prueba para variaciones en la forma de la distribución $\Delta\mu$.

3.2. Estudio de simulación

En esta sección se realiza el diseño de varios escenarios aplicando simulaciones Monte Carlo bajo una distribución F para poder estimar la potencia de los gráficos de control para cada uno de los escenarios de variación de tamaños de muestra y medidas de profundidad tanto para observaciones dependientes como independientes de tal forma que se pueda evaluar el desempeño de los mismos.

El estudio de simulación se realizará para las Fases I y II de los gráficos de control a ser evaluados mediante métodos de Monte Carlo y bootstrap, nótese que podemos considerar este método como una simulación Monte Carlo, sin embargo, la simulación bootstrap no aplica una simulación directa a los datos sino a la distribución empírica de la muestra. Para esto denotemos:

- MC: Simulación Monte Carlo directo de la población.
- Boot: Simulación Monte Carlo que permite estimar la distribución del estadístico a partir de muestras tomadas de la distribución empírica.

El rendimiento de métodos paramétricos o no paramétricos debe ser evaluado en base a qué tan rápido estos esquemas detectan una situación fuera de control. Es decir, que tan rápido el esquema genera una señal cuando un cambio actual ha ocurrido en el proceso, dado que los gráficos propuestos pueden detectar simultáneamente cambios en la media e incremento de escala del proceso que producen que el mismo no esté bajo control.

El rendimiento en control, de cada gráfico de Fase II, es evaluado usando las mismas distribuciones en control consideradas para las simulaciones de la Fase I. Estas evaluaciones se basan en el $ARL_0 = \frac{1}{\alpha}$ y la distribución de la longitud de corrida, para obtener un estimado de la distribución en control se usan simulaciones Monte Carlo.

Se generan muestras de referencia de la distribución en control y muestras de tamaño n de la misma distribución hasta que ocurra una primera señal fuera de control. El número de muestras generadas hasta que una señal fuera de control aparezca es la longitud de corrida, RL_i de la i -ésima simulación $i = 1, \dots, N$. ARL_0 es estimado como la longitud promedio de corridas de las N simulaciones, esto es,

$$\widehat{ARL}_0 = \frac{1}{N} \sum_{i=1}^n RL_i$$

El rendimiento fuera de control es medido con el ARL_1 . El cual es estimado ge-

nerando muestras aleatorias de la distribución en control denominados datos de referencia. Luego, muestras de tamaño n se generan de la distribución fuera de control hasta que el gráfico de una señal fuera de control. El procedimiento es repetido N veces y el ARL_1 es estimado como el número medio de corridas de las N simulaciones.

Una segunda medida de clasificación se basa en la probabilidad de error Tipo-II (β) o potencia de la prueba ($1 - \beta$). El error tipo II es el error que ocurre cuando el gráfico no da una señal fuera de control cuando el proceso está de hecho fuera de control. El ARL_1 debe ser lo más bajo posible para aumentar la probabilidad de poder identificar rápidamente eventos que lleven al proceso a estar fuera de control

La relación entre la potencia de la prueba y el ARL está dado por:

$$OOC\ ARL = \frac{1}{1 - \beta}$$

3.2.1. Evaluación de desempeño para gráfico de control de calibrado

La generación de observaciones se simula a partir de un proceso gaussiano con media μ y matriz de varianzas Σ , el algoritmo se basa en una factorización de matriz de covarianzas, la idea se basa en simular datos independientes y transformarlos linealmente de modo que el resultado tenga la covarianza deseada. Luego, si $Z \sim \mathcal{N}_d(\mathbf{0}, I_d)$ y $\Sigma = LL^t$ factorización de Cholesky, entonces:

$$\mathbf{X} = \mu + \Sigma\mathbf{Z} \sim \mathcal{N}_d(\mu, \Sigma)$$

Es decir, el algoritmo es el siguiente para el caso independiente:

1. Obtener la factorización de Cholesky $\Sigma = LL^t$
2. Simular $\mathbf{Z} = (Z_1, Z_2, \dots, Z_d)$ i.i.d $N_d(0, 1)$.
3. Hacer $\mathbf{X} = \mu + \Sigma\mathbf{Z}$
4. Repetir los pasos 2 y 3 de acuerdo al número de simulaciones deseadas.

Y para el caso dependiente:

$$\mathbf{X} = \mu + \Sigma\mathbf{Y}_{ij} + \rho\mathbf{Y}_{ij-1}$$

donde $\mathbf{Y}_{ij} = \Sigma\mathbf{Z}$

1. Obtener la factorización de Cholesky $\Sigma = LL^t$
2. Simular $\mathbf{Z} = (Z_1, Z_2, \dots, Z_d)$ i.i.d $N_d(0, 1)$.
3. Hacer $\mathbf{Y}_{ij} = \Sigma \mathbf{Z}$
4. Para cada $j = 1, \dots, p$ hacer:
 - $X_{i1} = Y_{i1}$
 - $X_{ij} = \rho Y_{ij-1} + Y_{ij} + \boldsymbol{\mu}$
5. Repetir los pasos 2,3 y 4 de acuerdo al número de simulaciones deseadas.

donde $\Sigma = e^{-\|x\|}$ y ρ el coeficiente de correlación de Pearson.

Para generar escenarios de simulación para cada uno de los eventos se considera cambios en el vector de medias $\boldsymbol{\mu}$ del proceso y para el caso dependiente variaciones en el coeficiente de correlación ρ . El gráfico 3.1 muestra la función de densidad de una distribución normal (línea roja) con media $\mu = 0$ y desviación estándar $\sigma = 1$, la línea azul punteada representa un cambio negativo en la forma ($\Delta\mu = -1$) y la línea verde entrecortada un cambio positivo ($\Delta\mu = 1$).

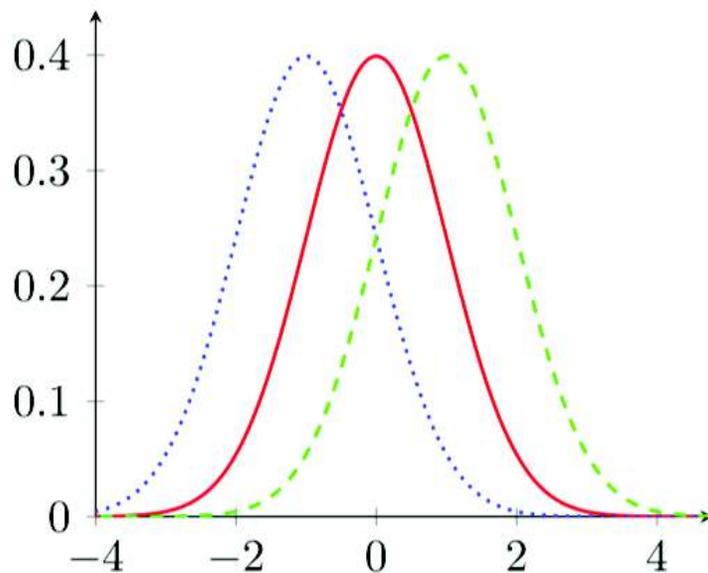


Figura 3.1: Cambio en la media del proceso

Comparación y rendimiento del gráfico de control propuesto para Fase I

A partir de la generación de muestras de referencia se compara el desempeño de los gráficos de control. Para cada muestra se calculan las medidas de profundidad

descritas en el capítulo 2 y se aplica los procedimientos bootstrap estándar y basado en recorte para la estimación del error tipo I cuando el proceso está en control y la potencia de la prueba cuando está fuera de control.

A partir de 500 remuestras ($B=500$), 1000 simulaciones Monte Carlo ($mc=1000$) y con procedimiento de recorte del 10 %, se aplica un bootstrap suavizado para el caso independiente con factor de suavización igual a $h = 0.01$ y un bootstrap estacionario para el caso dependiente para estimar el cuantil α que representa el límite de control inferior.

Esta metodología se puede resumir en los siguientes algoritmos para determinar tanto el límite de control inferior como la potencia de la prueba para la hipótesis alternativa. Así, para un vector de medias μ , una matriz de varianzas Σ y una medida de profundidad $D(\cdot)$ dadas y fijando $\alpha = 0.05$:

Caso independiente

1. Generar mc observaciones de la forma $\mathbf{X} = \mu + \Sigma\mathbf{Z}$
2. Para $i = 1, \dots, mc$ y $j = 1, \dots, B$ hacer:
 - $W_{ij} = D(\mathbf{X})$.
 - Extraer \hat{W}_{ij} , una muestra con reemplazo de tamaño n de W .
 - Calcular $R_{ij} = W_j + h\mathbf{Z}$ con $Z \sim \mathcal{N}(0, 1)$
 - Aplicar procedimiento de recorte de muestra.
3. Calcular LCI, el cuantil al nivel α para R_{ij} recortado.
4. Añadir una variación aislada, es decir un cambio en la forma de la distribución del proceso.
5. Calcular la potencia de la prueba $1 - \beta$ como el promedio de observaciones bajo LCI.

donde n es el número de filas de la matriz X y h es el coeficiente de suavizado del procedimiento bootstrap suavizado.

Caso dependiente

1. Generar mc observaciones de la forma $\mathbf{X} = \mu + \Sigma\mathbf{Y}_{ij} + \rho\mathbf{Y}_{ij-1}$ con $\mathbf{Y}_{ij} = \Sigma\mathbf{Z}$
2. Para $i = 1, \dots, mc$ hacer:

- $W = D(X)$.
 - Generar B muestras Bootstrap a partir del estadístico de profundidad aplicado a una serie temporal. Utilizando una longitud de bloque que sigue una distribución geométrica con media l .
 - Aplicar procedimiento de recorte de muestra.
3. Calcular LCI, el cuartil al nivel α para W_{ij} recortado.
 4. Añadir una variación aislada, es decir un cambio en la forma de la distribución del proceso.
 5. Calcular la potencia de la prueba $1 - \beta$ como el promedio de observaciones bajo LCI.

donde ρ el coeficiente de correlación.

El sobreestimar el error Tipo I ocasiona que los costos de producción aumenten, de tal forma que una buena gestión de calidad es indispensable si se desea tener un proceso óptimo. Se evalúa en el primer lugar bajo escenario de independencia para muestras de tamaño 15, 25, 50 y 100 bajo el algoritmo expuesto con anterioridad. En las tablas 3.1, 3.2, 3.3 y 3.4 se muestran los resultados de la estimación de la tasa de falsas alarmas. El porcentaje promedio de falsas señales fuera de control en la mayoría de los casos sobreestima al nivel de significancia fijado, notándose una lenta convergencia para el caso de la profundidad de Mahalanobis, e incluso aumentando el nivel de significancia al aumentar el tamaño de muestra para la profundidad de RP. Sin embargo, para muestras de tamaño 50 y 100 las medidas de profundidad Simplicial y Tukey dan valores más aproximados a α a pesar de que para el tamaño de muestra $n = 100$ la medida de profundidad de Mahalanobis subestima el nivel de significancia.

Cuadro 3.1: Fase I, Caso independiente para $n = 15$

El método de recorte de muestra para el caso bootstrap para la medida de profundidad de Tukey se desempeña de mejor manera que el resto de escenarios de simulación.

Muestra	Método		Mahalanobis	Tukey	RP	Simplicial
15	Estándar	MC	0.1623	0.1000	0.1875	0.1875
		Boot	0.1622	0.1000	0.2000	0.2000
	Trimmed	MC	0.1644	0.1000	0.2000	0.2000
		Boot	0.1644	0.1000	0.2000	0.2000

Cuadro 3.2: Fase I, Caso independiente para $n = 25$

El procedimiento estándar para el caso bootstrap y MC para la medida de profundidad de Tukey se desempeña de mejor manera, sin embargo, el resto de medidas de profundidad empiezan a tener una tendencia hacia α .

Muestra	Método		Mahalanobis	Tukey	RP	Simplicial
25	Estándar	MC	0.1546	0.0800	0.1638	0.1154
		Boot	0.1546	0.0800	0.1645	0.1200
	Trimmed	MC	0.1547	0.1000	0.1523	0.1200
		Boot	0.1547	0.1000	0.1524	0.1200

Cuadro 3.3: Fase I, Caso independiente para $n = 50$

El procedimiento estándar para el caso de simulación Monte Carlo presenta un valor cercano a α , podemos notar que el resto de medidas de profundidad presenta una convergencia lenta hacia α .

Muestra	Método		Mahalanobis	Tukey	RP	Simplicial
50	Estándar	MC	0.1485	0.0800	0.1885	0.0588
		Boot	0.1485	0.0800	0.1884	0.0600
	Trimmed	MC	0.1494	0.0800	0.1834	0.0600
		Boot	0.1494	0.0800	0.1835	0.0600

Cuadro 3.4: Fase I, Caso independiente para $n = 100$

El procedimiento estándar y boot para la medida simplicial subestima el nivel de confianza, notamos como la medida de Tukey converge lentamente a α .

Muestra	Método		Mahalanobis	Tukey	RP	Simplicial
100	Estándar	MC	0.1461	0.0750	0.2004	0.0297
		Boot	0.1460	0.0750	0.2004	0.0300
	Trimmed	MC	0.1462	0.0750	0.1973	0.0300
		Boot	0.1461	0.0750	0.1973	0.0300

Para evaluar la capacidad de gráfico de control para detectar un cambio en la forma de la media del proceso a través de la estimación de su potencia ($1 - \beta$) se añade una variación aislada, es decir, suponemos que el proceso experimentó una variación temporal en una observación de la muestra.

Evaluando el caso independiente resumido en las tablas 3.5, 3.6, 3.7 y 3.8 notamos que la medida de profundidad simplicial en general tiene un mejor desempeño para todos los tamaños de muestra y la tasa de detección correcta ($1 - \beta$) es más alta incluso ante pequeños cambios en la forma, con capacidad de detección alta ante tamaños de muestra pequeños y en general para todos los tamaños de muestra, sin embargo al aumentar el tamaño de muestra su potencia disminuye ante pequeñas variaciones. Por otro lado, el desempeño es similar en todos los escenarios donde se obtiene una mejor potencia para las profundidad de Mahalanobis y RP conforme aumenta el tamaño de muestra tanto para el método estándar como el recortado con excepción en el caso donde se usa la profundidad de Tukey, donde se tiene un menor rendimiento del gráfico de control para detectar correctas señales fuera de control. Sin embargo, cabe destacar que la medida de profundidad que arroja los mejores resultados es la simplicial.

Cuadro 3.5: Estimación de potencia de la prueba para $n = 15$ para caso independiente

El procedimiento recortado presenta mejores resultados en términos de convergencia para todas las medidas de profundidad destacando la medida Simplicial que a cambios en la forma del proceso converge a 1.

Muestra	Depth	Método	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$
15	Mahalanobis	Estándar	0.0520	0.3420	0.722	0.92
		Trimmed	0.0640	0.3490	0.731	0.927
	Tukey	Estándar	0.051	0.087	0.123	0.142
		Trimmed	0.055	0.094	0.128	0.146
	RP	Estándar	0.058	0.245	0.548	0.797
		Trimmed	0.104	0.309	0.582	0.799
	Simplicial	Estándar	0.555	0.836	0.976	0.999
		Trimmed	0.564	0.853	0.976	0.999

Cuadro 3.6: Estimación de potencia de la prueba para $n = 25$ para caso independiente

El procedimiento recortado presenta mejores resultados en términos de convergencia para todas las medidas de profundidad destacando la medida Simplicial y de Mahalanobis que a cambios en la magnitud del proceso converge a 1, la medida de RP también presenta una buena convergencia.

Muestra	Depth	Método	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$
25	Mahalanobis	Estándar	0.0860	0.3990	0.778	0.952
		Trimmed	0.1010	0.4160	0.791	0.953
	Tukey	Estándar	0.055	0.098	0.137	0.158
		Trimmed	0.063	0.106	0.144	0.165
	RP	Estándar	0.084	0.336	0.68	0.897
		Trimmed	0.112	0.37	0.681	0.91
	Simplicial	Estándar	0.409	0.761	0.942	0.997
		Trimmed	0.438	0.774	0.948	0.997

Cuadro 3.7: Estimación de potencia de la prueba para $n = 50$ para caso independiente

Se verifica la consistencia de los gráficos dado que al aumentar el tamaño de muestra aumenta la potencia de la prueba.

Muestra	Depth	Método	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$
50	Mahalanobis	Estándar	0.1230	0.4300	0.817	0.965
		Trimmed	0.1250	0.4400	0.81	0.964
	Tukey	Estándar	0.067	0.117	0.165	0.188
		Trimmed	0.069	0.117	0.168	0.192
	RP	Estándar	0.116	0.392	0.762	0.948
		Trimmed	0.135	0.42	0.778	0.946
	Simplicial	Estándar	0.285	0.656	0.91	0.99
		Trimmed	0.302	0.668	0.912	0.991

Cuadro 3.8: Estimación de potencia de la prueba para $n = 100$ para caso independiente

Las medidas de profundidad de Mahalanobis, RP y Simplicial presentan los mejores resultados, sin embargo esta última mide mejor la capacidad ante variaciones pequeñas de forma.

Muestra	Depth	Método	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$
100	Mahalanobis	Estándar	0.1450	0.4620	0.824	0.968
		Trimmed	0.1460	0.4640	0.826	0.973
	Tukey	Estándar	0.08	0.149	0.208	0.233
		Trimmed	0.084	0.154	0.213	0.241
	RP	Estándar	0.132	0.444	0.805	0.96
		Trimmed	0.137	0.458	0.81	0.964
	Simplicial	Estándar	0.208	0.526	0.858	0.975
		Trimmed	0.219	0.549	0.86	0.979

Para el caso dependiente se obtienen resultados de la tasa de falsa alarma similares al caso independiente para diferentes valores de ρ , las tablas 3.9, 3.10, 3.11 y 3.12 sintetizan los resultados de las estimaciones del porcentaje medio de falsas alarmas, podemos notar que conforme aumenta el tamaño de muestra n los porcentajes de error de tipo I se acercan al valor nominal, en general las medidas de profundidad de Tukey y Simplicial arrojan los mejores resultados con la primera acercándose al valor nominal conforme aumenta el tamaño de la muestra sin importar los niveles de dependencia y método empleado. A pesar de que la profundidad simplicial arroja los mejores resultados, cuando el tamaño de muestra es pequeño ($n = 15$) este sobrestima el porcentaje de error, sin embargo conforme el tamaño de muestra aumenta, esta se acerca rápidamente al valor nominal, aunque cuando el tamaño de muestra es igual a ($n = 100$) subestima este valor, notamos además la lenta convergencia de la medida de Mahalanobis y similar al caso independiente al aumentar el tamaño de muestra la profundidad RP sobrestima el nivel de confianza propuesto.

Cuadro 3.9: Resultados para el caso dependiente de la estimación de la tasa de falsas alarmas para $n = 15$

El método bootstrap estándar para la medida de profundidad RP arroja los mejores resultados para un coeficiente de correlación igual a 0.7 y 0.9. La medida de profundidad de Tukey presenta también buenos resultados para procedimiento MC estándar.

Muestra	ρ	Método		Mahalanobis	Tukey	RP	Simplicial
15	0.5	Estándar	MC	0.1657	0.1000	0.1392	0.1875
			Boot	0.0876	0.1333	0.0875	0.2000
		Trimmed	MC	0.1672	0.1000	0.1169	0.2000
			Boot	0.1875	0.1333	0.1876	0.2000
	0.7	Estándar	MC	0.1680	0.1000	0.1406	0.1875
			Boot	0.0879	0.1333	0.0879	0.2000
		Trimmed	MC	0.1688	0.1000	0.1182	0.2000
			Boot	0.1873	0.1333	0.1875	0.2000
	0.9	Estándar	MC	0.1682	0.1000	0.1384	0.1875
			Boot	0.0878	0.1333	0.0876	0.2000
		Trimmed	MC	0.1693	0.1000	0.1153	0.2000
			Boot	0.1871	0.1333	0.1872	0.2000

Cuadro 3.10: Resultados para el caso dependiente de la estimación de la tasa de falsas alarmas para $n = 25$

El procedimiento estándar para el caso de simulación MC para la medida de profundidad de Tukey se desempeña de mejor manera, sin embargo, el resto de medidas de profundidad empiezan a tener una tendencia hacia α .

Muestra	ρ	Método		Mahalanobis	Tukey	RP	Simplicial
25	0.5	Estándar	MC	0.1559	0.0800	0.1656	0.1154
			Boot	0.1319	0.1000	0.1318	0.1200
		Trimmed	MC	0.1672	0.1000	0.1169	0.2000
			Boot	0.1875	0.1333	0.1876	0.2000
	0.7	Estándar	MC	0.1566	0.0800	0.1656	0.1154
			Boot	0.1315	0.1000	0.1315	0.1200
		Trimmed	MC	0.1581	0.1000	0.1551	0.1200
			Boot	0.1207	0.1000	0.1200	0.1200
	0.9	Estándar	MC	0.1564	0.0800	0.1663	0.1154
			Boot	0.1313	0.1000	0.1311	0.1200
		Trimmed	MC	0.1580	0.1000	0.1556	0.1200
			Boot	0.1200	0.1000	0.1201	0.1200

Cuadro 3.11: Resultados para el caso dependiente de la estimación de la tasa de falsas alarmas para $n = 50$

La medida simplicial comienza a arrojar mejores resultados en comparación con el resto de medidas, obteniéndose un valor cercano a α para la simulación MC estándar.

Muestra	ρ	Método		Mahalanobis	Tukey	RP	Simplicial
50	0.5	Estándar	MC	0.1503	0.0800	0.1892	0.0588
			Boot	0.1695	0.0900	0.1696	0.0600
	Trimmed	MC	0.1503	0.0800	0.1839	0.0600	
		Boot	0.1646	0.0900	0.1651	0.0600	
0.7	Estándar	Estándar	MC	0.1501	0.0800	0.1891	0.0588
			Boot	0.1695	0.0900	0.1695	0.0600
	Trimmed	MC	0.1501	0.0800	0.1842	0.0600	
		Boot	0.1646	0.0900	0.1652	0.0600	
0.9	Estándar	Estándar	MC	0.1499	0.0800	0.1889	0.0588
			Boot	0.1694	0.0900	0.1694	0.0600
	Trimmed	MC	0.1503	0.0800	0.1840	0.0600	
		Boot	0.1646	0.0900	0.1650	0.0600	

Cuadro 3.12: Resultados para el caso dependiente de la estimación de la tasa de falsas alarmas para $n = 100$

La medida simplicial subestima el valor nominal, observamos la convergencia de la medida de Tukey conforme aumenta el tamaño de la muestra.

Muestra	ρ	Método		Mahalanobis	Tukey	RP	Simplicial
100	0.5	Estándar	MC	0.1468	0.0750	0.2018	0.0297
			Boot	0.1900	0.0750	0.1900	0.0300
	Trimmed	MC	0.1470	0.0750	0.1988	0.0300	
		Boot	0.1875	0.0800	0.1876	0.0300	
0.7	Estándar	Estándar	MC	0.1467	0.0750	0.2015	0.0297
			Boot	0.1899	0.0750	0.1899	0.0300
	Trimmed	MC	0.1465	0.0750	0.1992	0.0300	
		Boot	0.1873	0.0800	0.1875	0.0300	
0.9	Estándar	Estándar	MC	0.1466	0.0750	0.2011	0.0297
			Boot	0.1895	0.0750	0.1895	0.0300
	Trimmed	MC	0.1468	0.0750	0.1985	0.0300	
		Boot	0.1871	0.0800	0.1872	0.0300	

Al evaluar la capacidad de gráfico de control para detectar un cambio en la forma de la media del proceso a través de la estimación de su potencia ($1 - \beta$) para el caso dependiente se presentan resultados sintetizados en las tablas 3.13, 3.14, 3.15 y 3.16 para diferentes valores de ρ que van desde 0.5 hasta 0.9. En general el procedimiento tiene un mejor desempeño para las medidas de profundidad de Mahalanobis cuya potencia aumenta conforme aumenta el tamaño de muestra obteniendo resultados similares tanto para el método estándar como para el método recortados al igual que con la medida Random Projection; y la medida Simplicial que presenta el mejor desempeño para detectar cambios en la forma de la media del proceso. Por otro lado, la medida de Tukey no presenta una buena capacidad para la estimación de la potencia dada su lenta convergencia, se comprueba la consistencia de los gráficos de control dado que al aumentar el tamaño de muestra aumenta la potencia de los mismos.

Cuadro 3.13: Estimación de potencia de la prueba para $n = 15$ para caso dependencia

El procedimiento recortado presenta mejores resultados en términos de convergencia para la medida simplicial, podemos notar que a menor coeficiente de correlación es mayor la potencia.

Muestra	ρ	Depth	Método	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$	
15	0.5	Mahalanobis	Estándar	0.010	0.035	0.070	0.142	
			Trimmed	0.000	0.000	0.007	0.052	
			Tukey	Estándar	0.095	0.150	0.187	0.199
				Trimmed	0.099	0.148	0.189	0.202
		RP	Estándar	0.007	0.035	0.104	0.253	
			Trimmed	0.013	0.053	0.154	0.331	
			Simplicial	Estándar	0.524	0.814	0.962	0.997
				Trimmed	0.149	0.514	0.815	0.960
	0.7	Mahalanobis	Estándar	0.000	0.000	0.006	0.030	
			Trimmed	0.000	0.000	0.006	0.037	
			Tukey	Estándar	0.071	0.119	0.151	0.166
				Trimmed	0.080	0.121	0.161	0.175
		RP	Estándar	0.003	0.021	0.085	0.221	
			Trimmed	0.012	0.043	0.120	0.278	
			Simplicial	Estándar	0.489	0.790	0.950	0.993
				Trimmed	0.114	0.452	0.761	0.940
	0.9	Mahalanobis	Estándar	0.000	0.000	0.004	0.023	
			Trimmed	0.000	0.000	0.005	0.024	
			Tukey	Estándar	0.050	0.094	0.120	0.134
				Trimmed	0.063	0.101	0.136	0.150
		RP	Estándar	0.001	0.010	0.059	0.144	
			Trimmed	0.006	0.023	0.094	0.212	
			Simplicial	Estándar	0.438	0.746	0.929	0.988
				Trimmed	0.076	0.357	0.698	0.891

Cuadro 3.14: Estimación de potencia de la prueba para $n = 25$ para caso dependencia

Vale la pena destacar que el procedimiento para la medida simplicial sigue siendo mejor pero solo para el caso estándar, el caso recortado presenta una potencia bastante menor, notamos además que la medida de Mahalanobis presenta mejor potencia a cambios de forma grandes y coeficiente de correlación pequeño

Muestra	ρ	Depth	Método	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$
25	0.5	Mahalanobis	Estándar	0.023	0.202	0.576	0.851
			Trimmed	0.024	0.204	0.580	0.863
		Tukey	Estándar	0.051	0.086	0.125	0.139
			Trimmed	0.051	0.085	0.121	0.137
	RP	Estándar	0.013	0.053	0.154	0.331	
		Trimmed	0.013	0.053	0.154	0.331	
		Simplicial	Estándar	0.372	0.712	0.920	0.995
			Trimmed	0.018	0.136	0.481	0.806
0.7	Mahalanobis	Estándar	0.018	0.133	0.481	0.796	
		Trimmed	0.016	0.133	0.486	0.799	
		Tukey	Estándar	0.041	0.068	0.096	0.111
			Trimmed	0.034	0.064	0.095	0.113
	RP	Estándar	0.011	0.053	0.214	0.518	
		Trimmed	0.014	0.069	0.279	0.542	
		Simplicial	Estándar	0.323	0.668	0.902	0.985
			Trimmed	0.009	0.086	0.385	0.727
0.9	Mahalanobis	Estándar	0.003	0.069	0.359	0.700	
		Trimmed	0.009	0.082	0.360	0.701	
		Tukey	Estándar	0.028	0.049	0.072	0.083
			Trimmed	0.027	0.051	0.078	0.092
	RP	Estándar	0.005	0.035	0.149	0.410	
		Trimmed	0.009	0.045	0.184	0.434	
		Simplicial	Estándar	0.260	0.612	0.854	0.974
			Trimmed	0.006	0.048	0.274	0.621

Cuadro 3.15: Estimación de potencia de la prueba para $n = 50$ para caso dependencia

La profundidad de Mahalanobis y la Simplicial presentan mejores resultados, notamos como ahora el procedimiento recortado presenta mejor rendimiento, de la misma forma que para muestras pequeñas la potencia disminuye conforme aumenta el coeficiente de correlación.

Muestra	ρ	Depth	Método	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$
50	0.5	Mahalanobis	Estándar	0.115	0.419	0.800	0.961
			Trimmed	0.125	0.430	0.803	0.961
		Tukey	Estándar	0.058	0.104	0.153	0.179
			Trimmed	0.056	0.104	0.155	0.188
		RP	Estándar	0.041	0.169	0.518	0.838
			Trimmed	0.047	0.179	0.538	0.855
		Simplicial	Estándar	0.233	0.568	0.854	0.978
			Trimmed	0.254	0.603	0.884	0.981
	0.7	Mahalanobis	Estándar	0.081	0.335	0.728	0.941
			Trimmed	0.081	0.339	0.730	0.935
		Tukey	Estándar	0.045	0.081	0.123	0.147
			Trimmed	0.045	0.080	0.117	0.151
		RP	Estándar	0.019	0.122	0.413	0.776
			Trimmed	0.032	0.132	0.438	0.783
		Simplicial	Estándar	0.185	0.498	0.817	0.966
			Trimmed	0.213	0.556	0.884	0.978
	0.9	Mahalanobis	Estándar	0.053	0.237	0.638	0.907
			Trimmed	0.053	0.245	0.637	0.904
		Tukey	Estándar	0.024	0.054	0.084	0.112
			Trimmed	0.027	0.058	0.082	0.110
		RP	Estándar	0.009	0.077	0.288	0.655
			Trimmed	0.018	0.090	0.317	0.687
		Simplicial	Estándar	0.125	0.424	0.741	0.939
			Trimmed	0.154	0.434	0.753	0.941

Cuadro 3.16: Estimación de potencia de la prueba para $n = 100$ para caso dependencia

Conforme aumenta el tamaño de muestra notamos la consistencia de los gráficos al aumentar el tamaño de muestra aumenta la potencia de la prueba. La profundidad de Tukey presenta lenta convergencia.

Muestra	ρ	Depth	Método	$\mu = 1$	$\mu = 2$	$\mu = 3$	$\mu = 4$
100	0.5	Mahalanobis	Estándar	0.178	0.521	0.858	0.981
			Trimmed	0.181	0.518	0.855	0.981
		Tukey	Estándar	0.059	0.113	0.172	0.204
			Trimmed	0.057	0.113	0.177	0.208
	RP	Estándar	0.049	0.228	0.648	0.906	
			Trimmed	0.058	0.250	0.657	0.912
		Simplicial	Estándar	0.140	0.435	0.792	0.959
			Trimmed	0.100	0.345	0.715	0.954
0.7	Mahalanobis	Estándar	0.127	0.427	0.812	0.961	
		Trimmed	0.134	0.438	0.810	0.963	
		Tukey	Estándar	0.043	0.088	0.135	0.170
			Trimmed	0.041	0.087	0.140	0.173
	RP	Estándar	0.032	0.161	0.545	0.859	
			Trimmed	0.037	0.190	0.549	0.871
		Simplicial	Estándar	0.112	0.358	0.736	0.933
			Trimmed	0.135	0.376	0.766	0.945
0.9	Mahalanobis	Estándar	0.071	0.314	0.723	0.941	
		Trimmed	0.076	0.326	0.731	0.941	
		Tukey	Estándar	0.025	0.054	0.091	0.121
			Trimmed	0.024	0.051	0.091	0.124
	RP	Estándar	0.019	0.102	0.398	0.777	
			Trimmed	0.022	0.115	0.399	0.782
		Simplicial	Estándar	0.065	0.263	0.634	0.895
			Trimmed	0.100	0.289	0.687	0.909

3.2.2. Evaluación de desempeño gráfico de monitoreo

Para la etapa de monitoreo se compara la potencia del gráfico de control mediante un estudio Monte Carlo basado en los gráficos de control generados en la Fase I para medidas de profundidad, con el gráfico de control de rangos cuando el proceso no está bajo control.

El escenario de pruebas cambia, de detectar puntos de cambio y variaciones aisladas en una secuencia de datos observados, a encontrar desviaciones distribucionales entre datos de Fase II recién recolectados y datos de referencia. Los métodos de Fase II prueban la hipótesis:

$$H_0 : F = G$$

$$H_a : F \neq G$$

De manera similar al esquema de la Fase I la generación de observaciones se basa en una factorización de matriz de covarianzas de las observaciones simuladas por Monte Carlo, para las cuales sean $X \sim F$ la muestra de referencia y $Y \sim G$ la muestra de monitoreo, se calcula la profundidad de Y basado en X , es decir, $D_F(Y, X)$ a partir de 1000 remuestras ($B=1000$) se aplica un bootstrap suavizado con factor de suavización $\gamma = 0.01$ para el caso independiente y un bootstrap estacionario con $l = 5$, la media de la distribución geométrica usada para generar la longitud aleatoria del bloque para estimar y comparar la potencia del gráfico de control para detectar señales de fuera de control cuando se asume un nivel de significancia de $\alpha = 0.05$. Además, se consideran escenarios en los cuales se varía un componente del vector de medias y otro en el que se varían todos sus componentes para las simulaciones. Para la simulación de los escenarios descritos consideramos el siguiente algoritmo:

Caso independiente:

1. Para cada $\Delta\mu$ y tamaño de muestra n hacer:
 - Generar mc observaciones de la forma $Y = \mu + \Sigma Z$, muestra de monitoreo con variación en la forma.
 - A partir de la muestra de calibrado X calcular $D_F(Y, X)$ para $B = 500$ muestras bootstrap.
2. Determinar la potencia de la prueba $1 - \beta$ comparando con el límite de control inferior (LCI) teórico, bootstrap y rank calculados en la Fase I.

Caso dependiente:

1. Para cada $\Delta\mu$, coeficiente de correlación ρ y tamaño de muestra n hacer:
 - Generar mc observaciones de la forma $Y = \mu + \Sigma U_{ij} + \rho U_{ij-1}$ con $U_{ij} = \Sigma Z$, muestra de monitoreo con variación en la forma.
 - A partir de la muestra de calibrado X calcular $D_F(Y, X)$ para $B = 500$ muestras bootstrap.
2. Determinar la potencia de la prueba $1 - \beta$ comparando con el límite de control inferior (LCI) teórico, bootstrap y rank calculados en la Fase I.

Con Σ la matriz de varianzas y covarianzas.

Los escenarios planteados en la metodología se emplearán para el cálculo de la potencia de la prueba basado en los límites de control teóricos (MC), bootstrap y rank determinados en la Fase I y medidas de profundidad de Mahalanobis, RP, Tukey y Simplicial. Este análisis se empleará para distintos tamaños de muestra n y variaciones en la forma de la distribución $\Delta\mu$.

En las tablas 3.17, 3.18, 3.19 y 3.20 se presentan para el caso de independencia los resultados de la simulación cuando se varía una componente del vector de medias y en las tablas 3.21, 3.22, 3.23 y 3.24 los resultados de la simulación cuando se varían todas las componentes del vector de medias para detectar un cambio en la magnitud de la media del proceso a través de la estimación de su potencia ($1 - \beta$). Cuando se varían todas las componentes como podemos ver en las tablas se tiene mayor potencia para todos los tamaños de muestra presentados.

A partir de los resultados presentados en las tablas podemos notar que debido al acercamiento asintótico del estadístico rank cuando el tamaño de muestra es pequeño ($n = 15$) el gráfico de control tiene poca potencia. Por otro lado, a medida que el tamaño de muestra aumenta la potencia del gráfico bootstrap es muy cercano al teórico para el estadístico rank y además a excepción de la medida de profundidad de Tukey que converge lentamente a medida que se desplaza la media del proceso, el resto de medidas tienen un desempeño similar superando la potencia de los gráficos rank, en la mayoría de los gráficos se verifica además la consistencia y claramente mientras más cambia la forma, mayor es la capacidad de detección de los gráficos de control.

Cuadro 3.17: Potencia del gráfico de control $1 - \beta$ variando un componente del vector de medias para $n = 15$ para escenario de independencia

Dado el acercamiento asintótico del estadístico rank, este presenta menor potencia comparado respecto a los procedimientos MC y boot, destaca la medida simplicial con mayor potencia antes cambios en la forma.

Muestra		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$
15	Mahalanobis	PR Liu	0.6386	0.3489	0.1520	0.0728	0.0500	0.1012	0.2391	0.5029
		PR MC	0.8150	0.5559	0.2998	0.1599	0.0500	0.2177	0.4296	0.7095
		PR Boot	0.8148	0.5556	0.2996	0.1598	0.0500	0.2175	0.4293	0.7093
Tukey		PR Liu	0.3725	0.3091	0.2042	0.1126	0.0661	0.1656	0.2694	0.3500
		PR MC	0.4393	0.3816	0.2683	0.1588	0.0661	0.2140	0.3300	0.4142
		PR Boot	0.4393	0.3816	0.2683	0.1588	0.0661	0.2140	0.3300	0.4142
RP		PR Liu	0.2164	0.0900	0.0312	0.0131	0.0500	0.0357	0.0813	0.1706
		PR MC	0.8466	0.5840	0.2945	0.1410	0.0500	0.2377	0.4377	0.6945
		PR Boot	0.8448	0.5810	0.2918	0.1395	0.0500	0.2360	0.4352	0.6921
Simplicial		PR Liu	0.9848	0.9224	0.7651	0.5450	0.4111	0.4177	0.5844	0.7872
		PR MC	0.9848	0.9224	0.7651	0.5450	0.4111	0.4177	0.5844	0.7872
		PR Boot	0.9848	0.9224	0.7651	0.5450	0.4111	0.4177	0.5844	0.7872

Cuadro 3.18: Potencia del gráfico de control $1 - \beta$ variando un componente del vector de medias para $n = 25$ para escenario de independencia

De manera similar al caso de $n = 15$ la potencia para los casos de las simulaciones MC y boot supera el de Liu al tener tamaño de muestra pequeño, destacan la profundidad de Mahalanobis y RP, sin embargo la medida Simplicial obtiene mejores resultados, aunque se puede notar que presenta la misma potencia que los gráficos de Liu.

Muestra		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$	
25	Mahalanobis	PR Liu	0.8817	0.6412	0.3319	0.1339	0.0500	0.2670	0.5645	0.8374	0.9652
		PR MC	0.8677	0.6153	0.3071	0.1180	0.0500	0.2437	0.5370	0.8203	0.9598
		PR Boot	0.8676	0.6152	0.3070	0.1180	0.0500	0.2436	0.5369	0.8202	0.9598
Tukey		PR Liu	0.3225	0.2631	0.1682	0.0845	0.0509	0.1037	0.1954	0.2798	0.3269
		PR MC	0.4417	0.4035	0.3096	0.1882	0.0509	0.1641	0.2741	0.3782	0.4324
		PR Boot	0.4417	0.4035	0.3096	0.1882	0.0509	0.1641	0.2741	0.3782	0.4324
RP		PR Liu	0.8075	0.5343	0.2480	0.0786	0.0500	0.0814	0.2560	0.5464	0.8175
		PR MC	0.9126	0.7101	0.4074	0.1651	0.0500	0.1739	0.4250	0.7265	0.9210
		PR Boot	0.9144	0.7140	0.4118	0.1679	0.0500	0.1769	0.4295	0.7304	0.9227
Simplicial		PR Liu	0.9741	0.8761	0.6574	0.4110	0.2777	0.3840	0.5775	0.7865	0.9282
		PR MC	0.9741	0.8761	0.6574	0.4110	0.2777	0.3840	0.5775	0.7865	0.9282
		PR Boot	0.9741	0.8761	0.6574	0.4110	0.2777	0.3840	0.5775	0.7865	0.9282

Cuadro 3.19: Potencia del gráfico de control $1 - \beta$ variando un componente del vector de medias para $n = 50$ para escenario de independencia

Al aumentar el tamaño de muestra, la potencia del gráfico rank se acerca al de la simulación MC y Boot mostrando la potencia de los gráficos propuesto para detectar cambios en la forma.

Muestra		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$	
50	Mahalanobis	PR Liu	0.7402	0.4359	0.1776	0.0580	0.0500	0.0854	0.2578	0.5569	0.8326
		PR MC	0.8288	0.5583	0.2681	0.1034	0.0500	0.1461	0.3688	0.6756	0.8986
		PR Boot	0.8289	0.5584	0.2682	0.1035	0.0500	0.1462	0.3689	0.6757	0.8987
Tukey		PR Liu	0.3463	0.3074	0.2238	0.1263	0.0529	0.0896	0.1699	0.2644	0.3278
		PR MC	0.3791	0.3406	0.2564	0.1539	0.0529	0.1108	0.1987	0.2998	0.3635
		PR Boot	0.3791	0.3406	0.2564	0.1539	0.0529	0.1108	0.1987	0.2998	0.3635
RP		PR Liu	0.7110	0.4237	0.1930	0.0868	0.0500	0.1613	0.3425	0.6044	0.8396
		PR MC	0.9287	0.7544	0.4898	0.2899	0.0500	0.4101	0.6592	0.8724	0.9716
		PR Boot	0.9286	0.7540	0.4893	0.2894	0.0500	0.4096	0.6587	0.8721	0.9715
Simplicial		PR Liu	0.9543	0.8078	0.5337	0.2714	0.1593	0.1899	0.3625	0.6241	0.8538
		PR MC	0.9543	0.8078	0.5337	0.2714	0.1593	0.1899	0.3625	0.6241	0.8538
		PR Boot	0.9543	0.8078	0.5337	0.2714	0.1593	0.1899	0.3625	0.6241	0.8538

Cuadro 3.20: Potencia del gráfico de control $1 - \beta$ variando un componente del vector de medias para $n = 100$ para escenario de independencia

Potencia muy similar comparando contra el gráfico rank, se verifica la consistencia de los gráficos propuestos, notamos la convergencia lenta de la medida de Tukey, sin embargo, el resto de medidas presenta buenos resultados.

Muestra		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$
100	Mahalanobis	PR Liu	0.9092	0.6956	0.3851	0.1537	0.0500	0.1497	0.3797	0.6900
		PR MC	0.9008	0.6779	0.3659	0.1416	0.0500	0.1376	0.3602	0.6721
		PR Boot	0.9008	0.6778	0.3658	0.1415	0.0500	0.1375	0.3601	0.6721
Tukey		PR Liu	0.3320	0.2906	0.2083	0.1174	0.0500	0.0932	0.1717	0.2611
		PR MC	0.3339	0.2950	0.2155	0.1253	0.0500	0.0988	0.1781	0.2668
		PR Boot	0.3339	0.2950	0.2155	0.1253	0.0500	0.0988	0.1781	0.2668
RP		PR Liu	0.8970	0.6704	0.3557	0.1320	0.0500	0.2101	0.4892	0.7822
		PR MC	0.9005	0.6772	0.3626	0.1356	0.0500	0.2153	0.4968	0.7880
		PR Boot	0.9006	0.6773	0.3628	0.1357	0.0500	0.2154	0.4969	0.7881
Simplicial		PR Liu	0.9539	0.8055	0.5261	0.2533	0.0890	0.1352	0.3023	0.5833
		PR MC	0.9539	0.8055	0.5261	0.2533	0.0890	0.1352	0.3023	0.5833
		PR Boot	0.9539	0.8055	0.5261	0.2533	0.0890	0.1352	0.3023	0.5833

Cuadro 3.21: Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 15$ para escenario de independencia

Resultado muy similar al caso de variación de una componente del vector de medias, sin embargo para este caso la potencia de los gráficos para todas las medidas de profundidad aumenta.

Muestra		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$
15	Mahalanobis	PR Liu	0.9109	0.6699	0.3331	0.1104	0.0500	0.1727	0.4676	0.7951
		PR MC	0.9610	0.8019	0.4881	0.2087	0.0500	0.3019	0.6321	0.8937
		PR Boot	0.9610	0.8017	0.4879	0.2085	0.0500	0.3017	0.6319	0.8936
Tukey		PR Liu	0.9562	0.7959	0.4712	0.1726	0.0661	0.2728	0.6135	0.8810
		PR MC	0.9770	0.8608	0.5639	0.2366	0.0661	0.3241	0.6645	0.9051
		PR Boot	0.9770	0.8608	0.5639	0.2366	0.0661	0.3241	0.6645	0.9051
RP		PR Liu	0.2873	0.1528	0.0736	0.0335	0.0500	0.0126	0.0276	0.0798
		PR MC	0.9130	0.7080	0.4350	0.2339	0.0500	0.1367	0.2893	0.6005
		PR Boot	0.9116	0.7053	0.4322	0.2321	0.0500	0.1352	0.2864	0.5972
Simplicial		PR Liu	0.9964	0.9617	0.8116	0.5451	0.4111	0.5368	0.8062	0.9964
		PR MC	0.9964	0.9617	0.8116	0.5451	0.4111	0.5368	0.8062	0.9964
		PR Boot	0.9964	0.9617	0.8116	0.5451	0.4111	0.5368	0.8062	0.9964

Cuadro 3.22: Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 25$ para escenario de independencia

Resultado muy similar al caso de variación de una componente del vector de medias, sin embargo para este caso la potencia de los gráficos para todas las medidas de profundidad aumenta. La medida simplicial sigue presentando mejores resultados aunque la de Tukey es cercana.

Muestra	$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$
25									
Mahalanobis	PR Liu	0.9708	0.8290	0.5152	0.2130	0.0500	0.2306	0.5402	0.8434
	PR MC	0.9659	0.8107	0.4871	0.1919	0.0500	0.2090	0.5125	0.8263
	PR Boot	0.9659	0.8106	0.4870	0.1918	0.0500	0.2089	0.5124	0.8262
Tukey	PR Liu	0.9484	0.7714	0.4365	0.1457	0.0509	0.1733	0.4746	0.7914
	PR MC	0.9804	0.8782	0.6011	0.2664	0.0509	0.2529	0.5766	0.8601
	PR Boot	0.9804	0.8782	0.6011	0.2664	0.0509	0.2529	0.5766	0.8601
RP	PR Liu	0.6892	0.3620	0.1300	0.0460	0.0500	0.0527	0.1646	0.4398
	PR MC	0.8707	0.6010	0.2910	0.1221	0.0500	0.1379	0.3455	0.6730
	PR Boot	0.8737	0.6067	0.2959	0.1247	0.0500	0.1410	0.3508	0.6782
Simplicial	PR Liu	0.9919	0.9304	0.7236	0.4293	0.2777	0.4859	0.7891	0.9574
	PR MC	0.9919	0.9304	0.7236	0.4293	0.2777	0.4859	0.7891	0.9574
	PR Boot	0.9919	0.9304	0.7236	0.4293	0.2777	0.4859	0.7891	0.9574

Cuadro 3.23: Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 50$ para escenario de independencia

Resultado muy similar al caso de variación de una componente del vector de medias, las simulaciones teóricas y boot superar a los gráficos rank.

Muestra		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$	
50	Mahalanobis	PR Liu	0.9443	0.7388	0.3823	0.1131	0.0500	0.1128	0.3855	0.7432	0.9455
		PR MC	0.9684	0.8170	0.4873	0.1758	0.0500	0.1765	0.4918	0.8212	0.9692
		PR Boot	0.9684	0.8170	0.4874	0.1759	0.0500	0.1766	0.4919	0.8212	0.9692
Tukey		PR Liu	0.9742	0.8529	0.5525	0.2176	0.0529	0.1506	0.4386	0.7705	0.9467
		PR MC	0.9802	0.8787	0.6019	0.2594	0.0529	0.1771	0.4783	0.8004	0.9575
		PR Boot	0.9802	0.8787	0.6019	0.2594	0.0529	0.1771	0.4783	0.8004	0.9575
RP		PR Liu	0.8053	0.5277	0.2714	0.1321	0.0500	0.0955	0.2287	0.5251	0.8287
		PR MC	0.9700	0.8478	0.6016	0.3640	0.0500	0.3435	0.6036	0.8651	0.9776
		PR Boot	0.9698	0.8474	0.6011	0.3635	0.0500	0.3430	0.6031	0.8647	0.9775
Simplicial		PR Liu	0.9895	0.9118	0.6586	0.3123	0.1593	0.3093	0.6551	0.9098	0.9890
		PR MC	0.9895	0.9118	0.6586	0.3123	0.1593	0.3093	0.6551	0.9098	0.9890
		PR Boot	0.9895	0.9118	0.6586	0.3123	0.1593	0.3093	0.6551	0.9098	0.9890

Cuadro 3.24: Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 100$ para escenario de independencia

Resultado muy similar al caso de variación de una componente del vector de medias, las simulaciones teóricas y boot superar a los gráficos rank.

Muestra		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$	
100	Mahalanobis	PR Liu	0.9724	0.8335	0.5159	0.1997	0.0500	0.1661	0.4601	0.7952	0.9612
		PR MC	0.9692	0.8211	0.4965	0.1859	0.0500	0.1536	0.4410	0.7812	0.9570
		PR Boot	0.9691	0.8211	0.4964	0.1858	0.0500	0.1535	0.4409	0.7812	0.9570
Tukey		PR Liu	0.9726	0.8467	0.5409	0.2083	0.0500	0.1640	0.4659	0.7935	0.9557
		PR MC	0.9742	0.8537	0.5553	0.2207	0.0500	0.1712	0.4757	0.8001	0.9578
		PR Boot	0.9742	0.8537	0.5553	0.2207	0.0500	0.1712	0.4757	0.8001	0.9578
RP		PR Liu	0.8854	0.6234	0.3041	0.1233	0.0500	0.1450	0.3637	0.6920	0.9193
		PR MC	0.8910	0.6330	0.3124	0.1275	0.0500	0.1496	0.3723	0.7009	0.9234
		PR Boot	0.8911	0.6331	0.3125	0.1276	0.0500	0.1497	0.3725	0.7010	0.9235
Simplicial		PR Liu	0.9881	0.9045	0.6431	0.2957	0.0890	0.2033	0.5083	0.8255	0.9691
		PR MC	0.9881	0.9045	0.6431	0.2957	0.0890	0.2033	0.5083	0.8255	0.9691
		PR Boot	0.9881	0.9045	0.6431	0.2957	0.0890	0.2033	0.5083	0.8255	0.9691

Para el caso dependiente se presentan los resultados de detectar un desplazamiento en la media del proceso en las tablas 3.25, 3.26, 3.27 y 3.28 cuando varía una componente del vector de medias y en las tablas 3.29, 3.30, 3.31 y 3.32 cuando varían todos los componentes del vector de medias, como en el caso independiente al variar ambos componentes se tiene una mejor potencia en los gráficos de control. Dado que el gráfico rank propuesto por Liu asume independencia los resultados que arroja no son consistentes, por otro lado los resultados del método bootstrap estacionario son aceptables dado que, cuando aumenta el tamaño de muestra este se acerca al método Monte Carlo que en este caso se asumiría como el teórico. Además, podemos notar que a pesar de que el porcentaje de rechazo presenta una pequeña variación al aumentar la tasa de dependencia, en general se tiene un desempeño similar al utilizar cualquier medida de profundidad para los diferentes valores de ρ , aunque se destaca que mientras aumenta la correlación entre las variables, es decir, aumenta el coeficiente de correlación el porcentaje de rechazo disminuye para la profundidad simplicial y de Mahalanobis y aumenta para la profundidad de Tukey; respecto a la profundidad RP varía de acuerdo al tamaño de muestra. En resumen, la medida de profundidad Simplicial arroja los mejores resultados tanto para el caso dependiente como independiente mientras aumenta el tamaño de la muestra.

Cuadro 3.25: Potencia del gráfico de control $1 - \beta$ variando una componente del vector de medias para $n = 15$ para escenario de dependencia

Al aumentar el coeficiente de correlación la potencia aumenta en la mayoría de los casos, la profundidad Simplicial como en el caso independiente sigue presentando mejores resultados.

ρ	Depth		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$
0.5	Mahalanobis	PR Liu	0.7452	0.5054	0.2968	0.1886	0.0500	0.2724	0.4696	0.7134	0.8950
		PR MC	0.7958	0.5705	0.3534	0.2313	0.0500	0.3273	0.5354	0.7675	0.9225
		PR Boot	0.4278	0.2110	0.0969	0.0559	0.0500	0.0842	0.1842	0.3873	0.6536
	Tukey	PR Liu	0.6960	0.5786	0.4247	0.3228	0.0723	0.3742	0.4771	0.5996	0.6975
		PR MC	0.6960	0.5786	0.4247	0.3228	0.0723	0.3742	0.4771	0.5996	0.6975
		PR Boot	0.7022	0.5889	0.4366	0.3392	0.0723	0.4310	0.5435	0.6480	0.7185
	RP	PR Liu	0.6731	0.4295	0.2037	0.0700	0.0500	0.0776	0.2212	0.4520	0.6942
		PR MC	0.9685	0.8778	0.6884	0.4655	0.0500	0.4837	0.7093	0.8905	0.9729
		PR Boot	0.8424	0.6346	0.3722	0.1664	0.0500	0.1807	0.3963	0.6595	0.8583
Simplicial	PR Liu	0.9562	0.8512	0.6690	0.4905	0.4323	0.5007	0.6726	0.8477	0.9530	
	PR MC	0.9562	0.8512	0.6690	0.4905	0.4323	0.5007	0.6726	0.8477	0.9530	
	PR Boot	0.9562	0.8512	0.6690	0.4905	0.4323	0.5007	0.6726	0.8477	0.9530	
0.7	Mahalanobis	PR Liu	0.7476	0.5346	0.3445	0.2390	0.0500	0.3224	0.5036	0.7196	0.8858
		PR MC	0.7941	0.5944	0.3998	0.2846	0.0500	0.3768	0.5642	0.7692	0.9128
		PR Boot	0.4571	0.2527	0.1323	0.0836	0.0500	0.1191	0.2266	0.4212	0.6612
	Tukey	PR Liu	0.7059	0.5956	0.4598	0.3671	0.0723	0.4121	0.5104	0.6218	0.7116
		PR MC	0.7059	0.5956	0.4598	0.3671	0.0723	0.4121	0.5104	0.6218	0.7116
		PR Boot	0.7117	0.6047	0.4715	0.3850	0.0723	0.4644	0.5697	0.6665	0.7334
	RP	PR Liu	0.6626	0.4330	0.2206	0.0874	0.0500	0.0958	0.2380	0.4551	0.6830
		PR MC	0.9596	0.8669	0.6947	0.5060	0.0500	0.5208	0.7130	0.8790	0.9643
		PR Boot	0.8273	0.6288	0.3865	0.1956	0.0500	0.2097	0.4097	0.6516	0.8431
Simplicial	PR Liu	0.9487	0.8495	0.6907	0.5400	0.4323	0.5475	0.6952	0.8476	0.9462	
	PR MC	0.9487	0.8495	0.6907	0.5400	0.4323	0.5475	0.6952	0.8476	0.9462	
	PR Boot	0.9487	0.8495	0.6907	0.5400	0.4323	0.5475	0.6952	0.8476	0.9462	
0.9	Mahalanobis	PR Liu	0.7516	0.5664	0.3980	0.2997	0.0500	0.3786	0.5402	0.7276	0.8768
		PR MC	0.7941	0.6209	0.4514	0.3476	0.0500	0.4317	0.5951	0.7724	0.9034
		PR Boot	0.4889	0.3003	0.1779	0.1227	0.0500	0.1645	0.2764	0.4576	0.6706
	Tukey	PR Liu	0.7157	0.6150	0.4973	0.4161	0.0723	0.4538	0.5445	0.6447	0.7265
		PR MC	0.7157	0.6150	0.4973	0.4161	0.0723	0.4538	0.5445	0.6447	0.7265
		PR Boot	0.7211	0.6234	0.5091	0.4349	0.0723	0.5013	0.5967	0.6854	0.7482
	RP	PR Liu	0.6517	0.4388	0.2408	0.1113	0.0500	0.1203	0.2576	0.4597	0.6712
		PR MC	0.9493	0.8573	0.7066	0.5521	0.0500	0.5639	0.7220	0.8687	0.9544
		PR Boot	0.8111	0.6243	0.4058	0.2330	0.0500	0.2463	0.4270	0.6453	0.8264
Simplicial	PR Liu	0.9411	0.8491	0.7155	0.5926	0.4323	0.5979	0.7197	0.8490	0.9393	
	PR MC	0.9411	0.8491	0.7155	0.5926	0.4323	0.5979	0.7197	0.8490	0.9393	
	PR Boot	0.9411	0.8491	0.7155	0.5926	0.4323	0.5979	0.7197	0.8490	0.9393	

Cuadro 3.26: Potencia del gráfico de control $1 - \beta$ variando una componente del vector de medias para $n = 25$ para escenario de dependencia

Resultado similar al caso independiente, dado que los gráficos propuestos por Liu están diseñados para el caso independiente sus resultados no son muy verídicos.

ρ	Depth		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$
0.5	Mahalanobis	PR Liu	0.7154	0.4484	0.2232	0.1183	0.0500	0.2773	0.5277	0.7825	0.9345
		PR MC	0.7976	0.5548	0.3131	0.1838	0.0500	0.3750	0.6315	0.8520	0.9618
		PR Boot	0.7305	0.4665	0.2374	0.1280	0.0500	0.2932	0.5457	0.7956	0.9400
	Tukey	PR Liu	0.5715	0.4895	0.3773	0.3040	0.0546	0.3789	0.4598	0.5327	0.5829
		PR MC	0.3870	0.3161	0.2338	0.1870	0.0546	0.2528	0.3262	0.3864	0.4194
		PR Boot	0.4441	0.3690	0.2790	0.2329	0.0546	0.3207	0.3861	0.4361	0.4683
	RP	PR Liu	0.5302	0.2711	0.1016	0.0331	0.0500	0.0335	0.1027	0.2730	0.5319
		PR MC	0.8881	0.6986	0.4460	0.2484	0.0500	0.2502	0.4483	0.7012	0.8889
		PR Boot	0.7332	0.4717	0.2312	0.0981	0.0500	0.0988	0.2333	0.4738	0.7347
Simplicial	PR Liu	0.9443	0.8216	0.6240	0.4467	0.3016	0.4828	0.6639	0.8444	0.9522	
	PR MC	0.9443	0.8216	0.6240	0.4467	0.3016	0.4828	0.6639	0.8444	0.9522	
	PR Boot	0.9443	0.8216	0.6240	0.4467	0.3016	0.4828	0.6639	0.8444	0.9522	
0.7	Mahalanobis	PR Liu	0.7151	0.4765	0.2689	0.1649	0.0500	0.3205	0.5478	0.7770	0.9227
		PR MC	0.7921	0.5748	0.3590	0.2392	0.0500	0.4156	0.6436	0.8427	0.9521
		PR Boot	0.7294	0.4935	0.2834	0.1762	0.0500	0.3362	0.5645	0.7894	0.9286
	Tukey	PR Liu	0.5934	0.5140	0.4133	0.3474	0.0546	0.4131	0.4918	0.5616	0.6099
		PR MC	0.4170	0.3470	0.2696	0.2251	0.0546	0.2889	0.3607	0.4200	0.4547
		PR Boot	0.4719	0.3982	0.3159	0.2730	0.0546	0.3529	0.4189	0.4701	0.5028
	RP	PR Liu	0.5322	0.2933	0.1271	0.0514	0.0500	0.0516	0.1282	0.2950	0.5342
		PR MC	0.8752	0.6992	0.4747	0.2963	0.0500	0.2985	0.4772	0.7015	0.8766
		PR Boot	0.7233	0.4848	0.2645	0.1320	0.0500	0.1328	0.2662	0.4874	0.7252
Simplicial	PR Liu	0.9361	0.8216	0.6504	0.4995	0.3016	0.5298	0.6861	0.8438	0.9451	
	PR MC	0.9361	0.8216	0.6504	0.4995	0.3016	0.5298	0.6861	0.8438	0.9451	
	PR Boot	0.9361	0.8216	0.6504	0.4995	0.3016	0.5298	0.6861	0.8438	0.9451	
0.9	Mahalanobis	PR Liu	0.7173	0.5078	0.3218	0.2242	0.0500	0.3696	0.5707	0.7725	0.9099
		PR MC	0.7882	0.5981	0.4115	0.3056	0.0500	0.4612	0.6582	0.8345	0.9408
		PR Boot	0.7305	0.5235	0.3368	0.2371	0.0500	0.3848	0.5860	0.7840	0.9160
	Tukey	PR Liu	0.6151	0.5404	0.4527	0.3955	0.0546	0.4510	0.5245	0.5908	0.6373
		PR MC	0.4485	0.3811	0.3105	0.2693	0.0546	0.3283	0.3965	0.4551	0.4912
		PR Boot	0.5003	0.4303	0.3570	0.3180	0.0546	0.3886	0.4530	0.5047	0.5380
	RP	PR Liu	0.5369	0.3194	0.1601	0.0786	0.0500	0.0791	0.1613	0.3215	0.5393
		PR MC	0.8626	0.7027	0.5082	0.3530	0.0500	0.3544	0.5109	0.7048	0.8644
		PR Boot	0.7159	0.5023	0.3038	0.1763	0.0500	0.1778	0.3054	0.5051	0.7179
Simplicial	PR Liu	0.9281	0.8235	0.6797	0.5560	0.3016	0.5800	0.7103	0.8449	0.9379	
	PR MC	0.9281	0.8235	0.6797	0.5560	0.3016	0.5800	0.7103	0.8449	0.9379	
	PR Boot	0.9281	0.8235	0.6797	0.5560	0.3016	0.5800	0.7103	0.8449	0.9379	

Cuadro 3.27: Potencia del gráfico de control $1 - \beta$ variando una componente del vector de medias para $n = 50$ para escenario de dependencia

Para este tamaño de muestra la potencia disminuye conforme aumenta la correlación entre las variables (Caso simplicial)

ρ	Depth		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$
0.5	Mahalanobis	PR Liu	0.8182	0.5804	0.3250	0.1664	0.0500	0.2404	0.4653	0.7280	0.9082
		PR MC	0.7692	0.5118	0.2649	0.1257	0.0500	0.1882	0.3963	0.6674	0.8764
		PR Boot	0.8175	0.5794	0.3240	0.1657	0.0500	0.2395	0.4642	0.7271	0.9078
	Tukey	PR Liu	0.4828	0.4079	0.2959	0.1956	0.0500	0.1965	0.2986	0.4118	0.4859
		PR MC	0.3463	0.2876	0.1996	0.1171	0.0500	0.1002	0.1718	0.2617	0.3319
		PR Boot	0.3553	0.2969	0.2080	0.1249	0.0500	0.1135	0.1896	0.2800	0.3464
	RP	PR Liu	0.8853	0.6818	0.3994	0.1665	0.0500	0.1294	0.3332	0.6177	0.8489
		PR MC	0.9645	0.8572	0.6317	0.3643	0.0500	0.3211	0.5774	0.8220	0.9519
		PR Boot	0.9453	0.8062	0.5528	0.2853	0.0500	0.2412	0.4903	0.7604	0.9261
	Simplicial	PR Liu	0.9316	0.7818	0.5415	0.3186	0.1762	0.3156	0.5408	0.7855	0.9345
		PR MC	0.9316	0.7818	0.5415	0.3186	0.1762	0.3156	0.5408	0.7855	0.9345
		PR Boot	0.9316	0.7818	0.5415	0.3186	0.1762	0.3156	0.5408	0.7855	0.9345
0.7	Mahalanobis	PR Liu	0.8099	0.5951	0.3658	0.2154	0.0500	0.2879	0.4936	0.7283	0.8963
		PR MC	0.7638	0.5321	0.3065	0.1695	0.0500	0.2337	0.4287	0.6720	0.8647
		PR Boot	0.8092	0.5942	0.3648	0.2147	0.0500	0.2870	0.4926	0.7275	0.8959
	Tukey	PR Liu	0.5078	0.4318	0.3253	0.2318	0.0500	0.2349	0.3322	0.4393	0.5129
		PR MC	0.3771	0.3142	0.2249	0.1426	0.0500	0.1253	0.1982	0.2888	0.3621
		PR Boot	0.3860	0.3235	0.2343	0.1527	0.0500	0.1419	0.2184	0.3092	0.3780
	RP	PR Liu	0.8667	0.6706	0.4102	0.1927	0.0500	0.1578	0.3515	0.6131	0.8317
		PR MC	0.9527	0.8404	0.6312	0.3958	0.0500	0.3609	0.5852	0.8089	0.9395
		PR Boot	0.9308	0.7900	0.5553	0.3159	0.0500	0.2783	0.5026	0.7486	0.9113
	Simplicial	PR Liu	0.9210	0.7791	0.5657	0.3690	0.1762	0.3667	0.5650	0.7819	0.9234
		PR MC	0.9210	0.7791	0.5657	0.3690	0.1762	0.3667	0.5650	0.7819	0.9234
		PR Boot	0.9210	0.7791	0.5657	0.3690	0.1762	0.3667	0.5650	0.7819	0.9234
0.9	Mahalanobis	PR Liu	0.8030	0.6130	0.4118	0.2758	0.0500	0.3432	0.5252	0.7302	0.8844
		PR MC	0.7597	0.5557	0.3538	0.2255	0.0500	0.2876	0.4652	0.6791	0.8533
		PR Boot	0.8024	0.6121	0.4109	0.2750	0.0500	0.3423	0.5243	0.7295	0.8840
	Tukey	PR Liu	0.5336	0.4577	0.3590	0.2743	0.0500	0.2798	0.3688	0.4684	0.5406
		PR MC	0.4082	0.3416	0.2532	0.1731	0.0500	0.1569	0.2281	0.3174	0.3928
		PR Boot	0.4173	0.3516	0.2642	0.1862	0.0500	0.1778	0.2516	0.3404	0.4110
	RP	PR Liu	0.8464	0.6597	0.4235	0.2252	0.0500	0.1939	0.3734	0.6100	0.8133
		PR MC	0.9385	0.8240	0.6348	0.4345	0.0500	0.4076	0.5982	0.7965	0.9249
		PR Boot	0.9139	0.7741	0.5614	0.3530	0.0500	0.3232	0.5188	0.7383	0.8944
	Simplicial	PR Liu	0.9096	0.7778	0.5938	0.4264	0.1762	0.4249	0.5930	0.7796	0.9114
		PR MC	0.9096	0.7778	0.5938	0.4264	0.1762	0.4249	0.5930	0.7796	0.9114
		PR Boot	0.9096	0.7778	0.5938	0.4264	0.1762	0.4249	0.5930	0.7796	0.9114

Cuadro 3.28: Potencia del gráfico de control $1 - \beta$ variando una componente del vector de medias para $n = 100$ para escenario de dependencia

La potencia del gráfico de Liu supera a la profundidad de Tukey, sin embargo por los resultados de las demás simulaciones notamos que esta medida no es muy potente, la medida simplicial sigue arrojando mejores resultados.

ρ	Depth		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$
0.5	Mahalanobis	PR Liu	0.9014	0.7140	0.4492	0.2319	0.0500	0.2283	0.4437	0.7099	0.8994
		PR MC	0.8439	0.6116	0.3389	0.1509	0.0500	0.1476	0.3331	0.6060	0.8412
		PR Boot	0.9121	0.7357	0.4753	0.2534	0.0500	0.2500	0.4703	0.7318	0.9103
	Tukey	PR Liu	0.4019	0.3475	0.2585	0.1687	0.0500	0.1452	0.2268	0.3234	0.3909
		PR MC	0.2856	0.2431	0.1774	0.1125	0.0500	0.0942	0.1525	0.2243	0.2770
		PR Boot	0.3023	0.2582	0.1892	0.1202	0.0500	0.1015	0.1632	0.2385	0.2930
	RP	PR Liu	0.8908	0.6872	0.4027	0.1709	0.0500	0.1184	0.3039	0.5859	0.8313
		PR MC	0.9408	0.7938	0.5365	0.2768	0.0500	0.2136	0.4384	0.7142	0.9042
		PR Boot	0.9223	0.7510	0.4786	0.2274	0.0500	0.1676	0.3783	0.6615	0.8765
Simplicial	PR Liu	0.9269	0.7660	0.5076	0.2655	0.0995	0.2382	0.4736	0.7463	0.9204	
	PR MC	0.9269	0.7660	0.5076	0.2655	0.0995	0.2382	0.4736	0.7463	0.9204	
	PR Boot	0.9269	0.7660	0.5076	0.2655	0.0995	0.2382	0.4736	0.7463	0.9204	
0.7	Mahalanobis	PR Liu	0.8891	0.7145	0.4782	0.2809	0.0500	0.2783	0.4740	0.7111	0.8873
		PR MC	0.8325	0.6197	0.3736	0.1946	0.0500	0.1914	0.3689	0.6154	0.8299
		PR Boot	0.8999	0.7346	0.5031	0.3033	0.0500	0.3007	0.4989	0.7316	0.8982
	Tukey	PR Liu	0.4338	0.3758	0.2875	0.2000	0.0500	0.1781	0.2576	0.3524	0.4219
		PR MC	0.3206	0.2736	0.2047	0.1384	0.0500	0.1199	0.1794	0.2534	0.3105
		PR Boot	0.3370	0.2887	0.2168	0.1471	0.0500	0.1281	0.1906	0.2676	0.3262
	RP	PR Liu	0.8739	0.6803	0.4213	0.2051	0.0500	0.1554	0.3336	0.5905	0.8174
		PR MC	0.9271	0.7819	0.5482	0.3163	0.0500	0.2608	0.4636	0.7107	0.8900
		PR Boot	0.9071	0.7408	0.4934	0.2650	0.0500	0.2108	0.4061	0.6611	0.8622
Simplicial	PR Liu	0.9150	0.7615	0.5291	0.3114	0.0995	0.2852	0.4953	0.7393	0.9062	
	PR MC	0.9150	0.7615	0.5291	0.3114	0.0995	0.2852	0.4953	0.7393	0.9062	
	PR Boot	0.9150	0.7615	0.5291	0.3114	0.0995	0.2852	0.4953	0.7393	0.9062	
0.9	Mahalanobis	PR Liu	0.8767	0.7171	0.5118	0.3388	0.0500	0.3371	0.5086	0.7148	0.8753
		PR MC	0.8216	0.6303	0.4127	0.2485	0.0500	0.2462	0.4091	0.6270	0.8198
		PR Boot	0.8875	0.7357	0.5349	0.3614	0.0500	0.3600	0.5317	0.7335	0.8861
	Tukey	PR Liu	0.4660	0.4050	0.3194	0.2368	0.0500	0.2178	0.2923	0.3829	0.4538
		PR MC	0.3574	0.3061	0.2359	0.1700	0.0500	0.1526	0.2110	0.2849	0.3456
		PR Boot	0.3736	0.3211	0.2484	0.1797	0.0500	0.1620	0.2226	0.2989	0.3612
	RP	PR Liu	0.8563	0.6754	0.4441	0.2483	0.0500	0.2035	0.3684	0.5975	0.8039
		PR MC	0.9120	0.7721	0.5643	0.3634	0.0500	0.3173	0.4941	0.7098	0.8755
		PR Boot	0.8905	0.7325	0.5125	0.3116	0.0500	0.2647	0.4390	0.6632	0.8475
Simplicial	PR Liu	0.9019	0.7580	0.5543	0.3654	0.0995	0.3421	0.5223	0.7348	0.8908	
	PR MC	0.9019	0.7580	0.5543	0.3654	0.0995	0.3421	0.5223	0.7348	0.8908	
	PR Boot	0.9019	0.7580	0.5543	0.3654	0.0995	0.3421	0.5223	0.7348	0.8908	

Cuadro 3.29: Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 15$ para escenario de dependencia

Se puede destacar que el gráfico rank presenta mejor potencia en un par de escenarios, pero como se había dicho al ser diseñado para datos independientes sus resultados no son confiables.

ρ	Depth		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$
0.5	Mahalanobis	PR Liu	0.9254	0.7363	0.4424	0.2134	0.0500	0.3651	0.6658	0.8932	0.9811
		PR MC	0.9416	0.7751	0.4929	0.2562	0.0500	0.4169	0.7120	0.9150	0.9863
		PR Boot	0.7936	0.5057	0.2233	0.0733	0.0500	0.1585	0.4149	0.7263	0.9238
	Tukey	PR Liu	0.9618	0.8362	0.5849	0.3440	0.0723	0.5224	0.7868	0.9426	0.9912
		PR MC	0.9618	0.8362	0.5849	0.3440	0.0723	0.5224	0.7868	0.9426	0.9912
		PR Boot	0.9621	0.8389	0.5963	0.3689	0.0723	0.5537	0.8086	0.9517	0.9932
	RP	PR Liu	0.6747	0.4041	0.1811	0.0631	0.0500	0.0732	0.2106	0.4548	0.7275
		PR MC	0.9840	0.9103	0.7186	0.4704	0.0500	0.5183	0.7716	0.9367	0.9904
		PR Boot	0.8770	0.6545	0.3697	0.1611	0.0500	0.1865	0.4227	0.7130	0.9104
Simplicial	PR Liu	0.9886	0.9280	0.7550	0.5200	0.4323	0.5633	0.7977	0.9457	0.9921	
	PR MC	0.9886	0.9280	0.7550	0.5200	0.4323	0.5633	0.7977	0.9457	0.9921	
	PR Boot	0.9886	0.9280	0.7550	0.5200	0.4323	0.5633	0.7977	0.9457	0.9921	
0.7	Mahalanobis	PR Liu	0.9118	0.7304	0.4665	0.2598	0.0500	0.3996	0.6681	0.8797	0.9732
		PR MC	0.9290	0.7676	0.5144	0.3047	0.0500	0.4505	0.7112	0.9021	0.9799
		PR Boot	0.7802	0.5147	0.2522	0.1011	0.0500	0.1905	0.4335	0.7181	0.9089
	Tukey	PR Liu	0.9467	0.8166	0.5888	0.3830	0.0723	0.5318	0.7692	0.9256	0.9851
		PR MC	0.9467	0.8166	0.5888	0.3830	0.0723	0.5318	0.7692	0.9256	0.9851
		PR Boot	0.9473	0.8204	0.6014	0.4085	0.0723	0.5632	0.7913	0.9356	0.9878
	RP	PR Liu	0.6712	0.4183	0.2037	0.0817	0.0500	0.0933	0.2343	0.4675	0.7208
		PR MC	0.9776	0.8995	0.7232	0.5105	0.0500	0.5514	0.7703	0.9259	0.9856
		PR Boot	0.8640	0.6543	0.3914	0.1927	0.0500	0.2191	0.4419	0.7085	0.8975
Simplicial	PR Liu	0.9834	0.9181	0.7599	0.5621	0.4323	0.5966	0.7960	0.9355	0.9878	
	PR MC	0.9834	0.9181	0.7599	0.5621	0.4323	0.5966	0.7960	0.9355	0.9878	
	PR Boot	0.9834	0.9181	0.7599	0.5621	0.4323	0.5966	0.7960	0.9355	0.9878	
0.9	Mahalanobis	PR Liu	0.8969	0.7268	0.4953	0.3159	0.0500	0.4398	0.6731	0.8660	0.9630
		PR MC	0.9150	0.7620	0.5412	0.3629	0.0500	0.4893	0.7136	0.8889	0.9711
		PR Boot	0.7671	0.5258	0.2859	0.1393	0.0500	0.2294	0.4542	0.7104	0.8923
	Tukey	PR Liu	0.9284	0.7978	0.5975	0.4265	0.0723	0.5476	0.7540	0.9060	0.9754
		PR MC	0.9284	0.7978	0.5975	0.4265	0.0723	0.5476	0.7540	0.9060	0.9754
		PR Boot	0.9295	0.8027	0.6112	0.4520	0.0723	0.5783	0.7759	0.9169	0.9791
	RP	PR Liu	0.6682	0.4352	0.2307	0.1068	0.0500	0.1202	0.2616	0.4813	0.7146
		PR MC	0.9694	0.8890	0.7314	0.5562	0.0500	0.5900	0.7721	0.9146	0.9789
		PR Boot	0.8506	0.6549	0.4166	0.2324	0.0500	0.2587	0.4641	0.7046	0.8829
Simplicial	PR Liu	0.9764	0.9084	0.7686	0.6080	0.4323	0.6347	0.7984	0.9249	0.9816	
	PR MC	0.9764	0.9084	0.7686	0.6080	0.4323	0.6347	0.7984	0.9249	0.9816	
	PR Boot	0.9764	0.9084	0.7686	0.6080	0.4323	0.6347	0.7984	0.9249	0.9816	

Cuadro 3.30: Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 25$ para escenario de dependencia

Resultado similar a la variación en una componente en forma, como en el caso independiente la potencia aumenta respecto a si se vería una componente.

ρ	Depth		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$
0.5	Mahalanobis	PR Liu	0.8951	0.6709	0.3716	0.1665	0.0500	0.2886	0.5829	0.8476	0.9688
		PR MC	0.9324	0.7531	0.4675	0.2412	0.0500	0.3847	0.6785	0.8987	0.9825
		PR Boot	0.9023	0.6857	0.3874	0.1780	0.0500	0.3043	0.5998	0.8575	0.9717
	Tukey	PR Liu	0.9636	0.8431	0.5976	0.3549	0.0546	0.5058	0.7673	0.9310	0.9874
		PR MC	0.9338	0.7655	0.4836	0.2422	0.0546	0.3812	0.6780	0.8954	0.9797
		PR Boot	0.9420	0.7865	0.5152	0.2820	0.0546	0.4570	0.7339	0.9170	0.9839
	RP	PR Liu	0.8024	0.5141	0.2199	0.0589	0.0500	0.0609	0.2263	0.5234	0.8084
		PR MC	0.9726	0.8614	0.6057	0.3080	0.0500	0.3148	0.6142	0.8659	0.9737
		PR Boot	0.9158	0.7092	0.3940	0.1445	0.0500	0.1490	0.4027	0.7166	0.9189
Simplicial	PR Liu	0.9852	0.9133	0.7253	0.4861	0.3016	0.5361	0.7805	0.9398	0.9910	
	PR MC	0.9852	0.9133	0.7253	0.4861	0.3016	0.5361	0.7805	0.9398	0.9910	
	PR Boot	0.9852	0.9133	0.7253	0.4861	0.3016	0.5361	0.7805	0.9398	0.9910	
0.7	Mahalanobis	PR Liu	0.8814	0.6722	0.4024	0.2110	0.0500	0.3292	0.5950	0.8363	0.9590
		PR MC	0.9202	0.7490	0.4938	0.2909	0.0500	0.4231	0.6833	0.8868	0.9754
		PR Boot	0.8888	0.6861	0.4178	0.2236	0.0500	0.3447	0.6104	0.8459	0.9623
	Tukey	PR Liu	0.9484	0.8217	0.5979	0.3903	0.0546	0.5155	0.7483	0.9115	0.9794
		PR MC	0.9116	0.7403	0.4839	0.2729	0.0546	0.3971	0.6625	0.8724	0.9688
		PR Boot	0.9220	0.7628	0.5169	0.3152	0.0546	0.4655	0.7138	0.8950	0.9744
	RP	PR Liu	0.7897	0.5227	0.2468	0.0804	0.0500	0.0827	0.2531	0.5315	0.7956
		PR MC	0.9633	0.8482	0.6120	0.3467	0.0500	0.3530	0.6201	0.8528	0.9644
		PR Boot	0.9014	0.7031	0.4146	0.1772	0.0500	0.1818	0.4231	0.7101	0.9048
Simplicial	PR Liu	0.9790	0.9034	0.7325	0.5302	0.3016	0.5712	0.7798	0.9292	0.9864	
	PR MC	0.9790	0.9034	0.7325	0.5302	0.3016	0.5712	0.7798	0.9292	0.9864	
	PR Boot	0.9790	0.9034	0.7325	0.5302	0.3016	0.5712	0.7798	0.9292	0.9864	
0.9	Mahalanobis	PR Liu	0.8677	0.6755	0.4380	0.2664	0.0500	0.3760	0.6098	0.8255	0.9472
		PR MC	0.9067	0.7466	0.5248	0.3496	0.0500	0.4666	0.6907	0.8747	0.9662
		PR Boot	0.8750	0.6883	0.4527	0.2799	0.0500	0.3913	0.6240	0.8346	0.9510
	Tukey	PR Liu	0.9294	0.8006	0.6020	0.4298	0.0546	0.5309	0.7320	0.8892	0.9673
		PR MC	0.8853	0.7157	0.4887	0.3090	0.0546	0.4167	0.6477	0.8463	0.9524
		PR Boot	0.8979	0.7398	0.5226	0.3531	0.0546	0.4781	0.6951	0.8696	0.9599
	RP	PR Liu	0.7767	0.5322	0.2772	0.1091	0.0500	0.1119	0.2831	0.5404	0.7823
		PR MC	0.9519	0.8352	0.6217	0.3931	0.0500	0.3986	0.6292	0.8395	0.9531
		PR Boot	0.8857	0.6978	0.4378	0.2178	0.0500	0.2223	0.4454	0.7045	0.8892
Simplicial	PR Liu	0.9711	0.8942	0.7436	0.5784	0.3016	0.6112	0.7832	0.9181	0.9797	
	PR MC	0.9711	0.8942	0.7436	0.5784	0.3016	0.6112	0.7832	0.9181	0.9797	
	PR Boot	0.9711	0.8942	0.7436	0.5784	0.3016	0.6112	0.7832	0.9181	0.9797	

Cuadro 3.31: Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 50$ para escenario de dependencia

Todas las medidas de profundidad presentan buenos resultados, notamos que en este caso la medida de Tukey tiene una potencia cercana al resto de medidas.

ρ	Depth		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$
0.5	Mahalanobis	PR Liu	0.9584	0.8143	0.5299	0.2479	0.0500	0.2500	0.5370	0.8204	0.9608
		PR MC	0.9434	0.7720	0.4694	0.1993	0.0500	0.2002	0.4753	0.7784	0.9461
		PR Boot	0.9582	0.8137	0.5290	0.2472	0.0500	0.2492	0.5361	0.8198	0.9605
	Tukey	PR Liu	0.9640	0.8378	0.5671	0.2731	0.0500	0.2897	0.5866	0.8471	0.9662
		PR MC	0.9396	0.7732	0.4736	0.1924	0.0500	0.1565	0.4094	0.7153	0.9126
		PR Boot	0.9397	0.7735	0.4749	0.1955	0.0500	0.1859	0.4643	0.7694	0.9402
	RP	PR Liu	0.8582	0.6346	0.3604	0.1547	0.0500	0.1133	0.2761	0.5308	0.7817
		PR MC	0.9685	0.8613	0.6339	0.3696	0.0500	0.2910	0.5190	0.7762	0.9350
		PR Boot	0.9434	0.7948	0.5365	0.2813	0.0500	0.2157	0.4277	0.6975	0.8955
Simplicial	PR Liu	0.9822	0.8945	0.6647	0.3708	0.1762	0.3471	0.6346	0.8771	0.9773	
	PR MC	0.9822	0.8945	0.6647	0.3708	0.1762	0.3471	0.6346	0.8771	0.9773	
	PR Boot	0.9822	0.8945	0.6647	0.3708	0.1762	0.3471	0.6346	0.8771	0.9773	
0.7	Mahalanobis	PR Liu	0.9468	0.8032	0.5444	0.2897	0.0500	0.2934	0.5523	0.8094	0.9495
		PR MC	0.9304	0.7624	0.4879	0.2394	0.0500	0.2415	0.4946	0.7690	0.9335
		PR Boot	0.9466	0.8026	0.5436	0.2889	0.0500	0.2926	0.5514	0.8088	0.9493
	Tukey	PR Liu	0.9477	0.8113	0.5588	0.2990	0.0500	0.3136	0.5770	0.8218	0.9508
		PR MC	0.9175	0.7434	0.4647	0.2104	0.0500	0.1755	0.4048	0.6861	0.8863
		PR Boot	0.9176	0.7442	0.4670	0.2154	0.0500	0.2079	0.4595	0.7416	0.9187
	RP	PR Liu	0.8440	0.6331	0.3799	0.1834	0.0500	0.1412	0.2984	0.5354	0.7694
		PR MC	0.9598	0.8510	0.6419	0.4061	0.0500	0.3334	0.5366	0.7693	0.9233
		PR Boot	0.9321	0.7863	0.5490	0.3168	0.0500	0.2536	0.4480	0.6941	0.8822
Simplicial	PR Liu	0.9742	0.8805	0.6685	0.4121	0.1762	0.3920	0.6429	0.8643	0.9687	
	PR MC	0.9742	0.8805	0.6685	0.4121	0.1762	0.3920	0.6429	0.8643	0.9687	
	PR Boot	0.9742	0.8805	0.6685	0.4121	0.1762	0.3920	0.6429	0.8643	0.9687	
0.9	Mahalanobis	PR Liu	0.9333	0.7927	0.5628	0.3401	0.0500	0.3444	0.5708	0.7990	0.9364
		PR MC	0.9153	0.7544	0.5097	0.2884	0.0500	0.2916	0.5170	0.7609	0.9188
		PR Boot	0.9330	0.7921	0.5620	0.3392	0.0500	0.3435	0.5699	0.7985	0.9362
	Tukey	PR Liu	0.9261	0.7839	0.5535	0.3306	0.0500	0.3426	0.5700	0.7948	0.9304
		PR MC	0.8895	0.7118	0.4574	0.2321	0.0500	0.1997	0.4024	0.6564	0.8547
		PR Boot	0.8899	0.7134	0.4613	0.2399	0.0500	0.2348	0.4570	0.7122	0.8916
	RP	PR Liu	0.8299	0.6337	0.4033	0.2201	0.0500	0.1771	0.3252	0.5424	0.7579
		PR MC	0.9492	0.8417	0.6536	0.4492	0.0500	0.3834	0.5589	0.7648	0.9104
		PR Boot	0.9193	0.7792	0.5660	0.3593	0.0500	0.3003	0.4729	0.6930	0.8679
Simplicial	PR Liu	0.9640	0.8672	0.6764	0.4604	0.1762	0.4444	0.6557	0.8523	0.9580	
	PR MC	0.9640	0.8672	0.6764	0.4604	0.1762	0.4444	0.6557	0.8523	0.9580	
	PR Boot	0.9640	0.8672	0.6764	0.4604	0.1762	0.4444	0.6557	0.8523	0.9580	

Cuadro 3.32: Potencia del gráfico de control $1 - \beta$ variando todas las componentes del vector de medias para $n = 100$ para escenario de dependencia

Se verifica en ciertos casos la consistencia de los gráficos, en todos los casos la potencia aumenta al variar en forma las observaciones, sin embargo la mejor medida a ser utilizada por todos los resultados anteriores en la Simplicial.

ρ	Depth		$\mu = -3.2$	$\mu = -2.4$	$\mu = -1.6$	$\mu = -0.8$	$\mu = 0$	$\mu = 0.8$	$\mu = 1.6$	$\mu = 2.4$	$\mu = 3.2$
0.5	Mahalanobis	PR Liu	0.9584	0.8143	0.5299	0.2479	0.0500	0.2500	0.5370	0.8204	0.9608
		PR MC	0.9434	0.7720	0.4694	0.1993	0.0500	0.2002	0.4753	0.7784	0.9461
		PR Boot	0.9582	0.8137	0.5290	0.2472	0.0500	0.2492	0.5361	0.8198	0.9605
	Tukey	PR Liu	0.9640	0.8378	0.5671	0.2731	0.0500	0.2897	0.5866	0.8471	0.9662
		PR MC	0.9396	0.7732	0.4736	0.1924	0.0500	0.1565	0.4094	0.7153	0.9126
		PR Boot	0.9397	0.7735	0.4749	0.1955	0.0500	0.1859	0.4643	0.7694	0.9402
	RP	PR Liu	0.8582	0.6346	0.3604	0.1547	0.0500	0.1133	0.2761	0.5308	0.7817
		PR MC	0.9685	0.8613	0.6339	0.3696	0.0500	0.2910	0.5190	0.7762	0.9350
		PR Boot	0.9434	0.7948	0.5365	0.2813	0.0500	0.2157	0.4277	0.6975	0.8955
Simplicial	PR Liu	0.9822	0.8945	0.6647	0.3708	0.1762	0.3471	0.6346	0.8771	0.9773	
	PR MC	0.9822	0.8945	0.6647	0.3708	0.1762	0.3471	0.6346	0.8771	0.9773	
	PR Boot	0.9822	0.8945	0.6647	0.3708	0.1762	0.3471	0.6346	0.8771	0.9773	
0.7	Mahalanobis	PR Liu	0.9468	0.8032	0.5444	0.2897	0.0500	0.2934	0.5523	0.8094	0.9495
		PR MC	0.9304	0.7624	0.4879	0.2394	0.0500	0.2415	0.4946	0.7690	0.9335
		PR Boot	0.9466	0.8026	0.5436	0.2889	0.0500	0.2926	0.5514	0.8088	0.9493
	Tukey	PR Liu	0.9477	0.8113	0.5588	0.2990	0.0500	0.3136	0.5770	0.8218	0.9508
		PR MC	0.9175	0.7434	0.4647	0.2104	0.0500	0.1755	0.4048	0.6861	0.8863
		PR Boot	0.9176	0.7442	0.4670	0.2154	0.0500	0.2079	0.4595	0.7416	0.9187
	RP	PR Liu	0.8440	0.6331	0.3799	0.1834	0.0500	0.1412	0.2984	0.5354	0.7694
		PR MC	0.9598	0.8510	0.6419	0.4061	0.0500	0.3334	0.5366	0.7693	0.9233
		PR Boot	0.9321	0.7863	0.5490	0.3168	0.0500	0.2536	0.4480	0.6941	0.8822
Simplicial	PR Liu	0.9742	0.8805	0.6685	0.4121	0.1762	0.3920	0.6429	0.8643	0.9687	
	PR MC	0.9742	0.8805	0.6685	0.4121	0.1762	0.3920	0.6429	0.8643	0.9687	
	PR Boot	0.9742	0.8805	0.6685	0.4121	0.1762	0.3920	0.6429	0.8643	0.9687	
0.9	Mahalanobis	PR Liu	0.9333	0.7927	0.5628	0.3401	0.0500	0.3444	0.5708	0.7990	0.9364
		PR MC	0.9153	0.7544	0.5097	0.2884	0.0500	0.2916	0.5170	0.7609	0.9188
		PR Boot	0.9330	0.7921	0.5620	0.3392	0.0500	0.3435	0.5699	0.7985	0.9362
	Tukey	PR Liu	0.9261	0.7839	0.5535	0.3306	0.0500	0.3426	0.5700	0.7948	0.9304
		PR MC	0.8895	0.7118	0.4574	0.2321	0.0500	0.1997	0.4024	0.6564	0.8547
		PR Boot	0.8899	0.7134	0.4613	0.2399	0.0500	0.2348	0.4570	0.7122	0.8916
	RP	PR Liu	0.8299	0.6337	0.4033	0.2201	0.0500	0.1771	0.3252	0.5424	0.7579
		PR MC	0.9492	0.8417	0.6536	0.4492	0.0500	0.3834	0.5589	0.7648	0.9104
		PR Boot	0.9193	0.7792	0.5660	0.3593	0.0500	0.3003	0.4729	0.6930	0.8679
Simplicial	PR Liu	0.9640	0.8672	0.6764	0.4604	0.1762	0.4444	0.6557	0.8523	0.9580	
	PR MC	0.9640	0.8672	0.6764	0.4604	0.1762	0.4444	0.6557	0.8523	0.9580	
	PR Boot	0.9640	0.8672	0.6764	0.4604	0.1762	0.4444	0.6557	0.8523	0.9580	

3.3. Caso de estudio de extrusión de palitos de chupete

En esta sección se aplica la metodología desarrollada para el diseño de gráficos de control para las Fases I y II aplicado al caso de estudio de extrusión de palitos de chupetes fabricados en Confiteca C.A. en la máquina troqueladora aquarius RLM6-B y CARUGIL.

Un palito de chupete óptimo (Figura 3.2) es recto, tiene corte de extremos recto, sin rebabas, hueco en el centro, debe ser rígido, no flexible, uniforme y sin defectos, además debe ser elaborado de material puro o virgen.



Figura 3.2: Palito de chupete

Estos se forman a partir de polipropileno rígido mediante un proceso denominado extrusión, el cual es un proceso industrial mecánico en donde se realiza una acción de moldeado del plástico, que por flujo continuo con presión y empuje, se lo hace pasar por un molde encargado de darle la forma deseada. El polímero fundido es forzado a pasar a través de un dado también llamado cabezal, por medio del empuje generado por la acción giratoria de un husillo (tornillo de Arquímedes) que gira concéntricamente en una cámara a temperaturas controladas llamada cañón, con una separación milimétrica entre ambos elementos. El material polimérico es alimentado por medio de una tolva en un extremo de la máquina y debido a la acción de empuje se funde, fluye y mezcla en el cañón y se obtiene por el otro lado con un perfil geométrico preestablecido.

El proceso parte al colocar el polipropileno en la tolva de alimentación el cual pasa al dosificador que regula la alimentación del material hacia la extrusora y así



(a) Polipropileno



(b) Mezcladora

Figura 3.3: Proceso mezclado de insumos

evitar variaciones de amperaje muy fuertes para posteriormente pasar al cañón donde se lleva a cabo la plastificación a altas temperaturas monitoreada por un tablero de control que regula la temperatura del cañón, temperatura del cabezal, velocidad del motor principal y velocidad del dosificador.

Una vez derretido el plástico pasa al cabezal donde el moldeador a partir de válvu-



(a) Tolva



(b) Panel de control

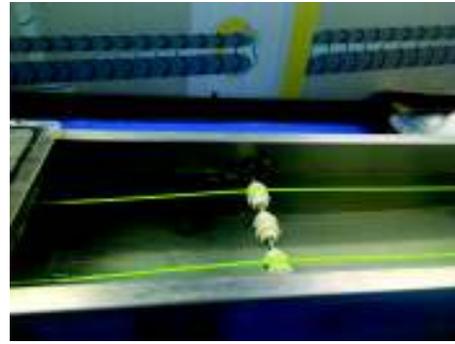
Figura 3.4: Componentes extrusora

las de aire forma el diámetro interno del cordón, dado que este sale a una temperatura promedio de 184 a 190 grados centígrados pasa a la tina de enfriamiento y de vacío donde regula la temperatura del cordón y la seca haciendo uso de aire. Cabe recalcar que por medio de la oruga se puede mantener una velocidad constante en el jalado o estirado del polipropileno para mantener el diámetro del cordón y controlar el espesor.

Finalmente, el operador afloja el cordón si se encuentra muy ancho y lo compri-



(a) Moldeador



(b) Tina de enfriamiento

Figura 3.5: Componentes extrusora II

me si está alargado en la zona previa a la cortadora donde se determina la longitud deseada del palito, posteriormente estos son pesados y empacados en sacos plásticos de polietileno y externo polipropileno, a fin de proteger y mantener su limpieza e integridad por efectos de transporte, manipuleo y almacenamiento.

Datos se recolectan durante tres meses de manufactura para distinto kilaje de pro-



(a) Cortadora



(b) Producto final

Figura 3.6: Componentes extrusora II

ducción, esta información es sintetizada en hojas proporcionadas por el departamento de calidad recabadas por los operadores de turno encargados de recolectar la información por hora de las medidas de calidad a ser evaluadas, estos datos se recogen cada 5 minutos y son sintetizados por hora como el promedio de las observaciones; y finalmente el dato que se recoge en el día es el promedio de las observaciones recolectadas por hora.

Tres medidas de calidad determinantes para la calidad final del producto se estu-

dian en esta aplicación las cuales son: diámetro interno, longitud y peso, para el análisis de la calidad global del palito el departamento de calidad usa índice de calidad calculado en función de si las características de calidad están en control individualmente. Actualmente esta metodología no está arrojando datos certeros acerca del verdadero estado de calidad del producto dado que a pesar de que el índice determina que un lote está en control, al realizar nuevamente pruebas para el armado junto con el caramelo este arroja resultados inconsistentes, más aún, al evaluar aleatoriamente muestras de producciones se obtienen resultados poco alentadores respecto a la gestión de calidad realizada.



Figura 3.7: Características de calidad

Dado que la longitud se define en el proceso de cortado y el peso se determina en base a la longitud del palito, calculando la correlación lineal entre estas variables obtenemos una correlación positiva de alrededor de 0.5, por tal motivo, asumimos que estas variables están relacionados y serán usadas para el caso de estudio de observaciones dependientes.

Por otro lado, en el caso independiente las variables de diámetro interno del palito y el peso se determinan en distintos estados del proceso productivo entonces podemos asumir que no están relacionadas, esta suposición la afirmamos realizando un test de correlación obteniendo un resultado muy cercano a cero. Los gráficos de control propuestos se comparan con los gráficos de control estándar (T^2 de Hotelling) para la Fase II en donde se contrasta la capacidad de detección de anomalías previamente detectadas por el departamento de calidad.

En base a la información descrita, para la Fase I se aplica el gráfico de control de profundidades para la estabilización del proceso y en la Fase II para monitoreo de las observaciones. Se consideran mediciones de los dos primeros meses para la fase de estabilización para lo cual se eliminan de los cálculos para estimar el límite de control aquellos puntos que son detectados como atípicos siempre y cuando obedezcan a una causa asignable, como se había comentado anteriormente estas causas se basan en observaciones emitidas por el departamento de calidad en donde se reportó influencias extrañas en el desempeño del proceso.

Caso Independiente

Para el caso de estudio de observaciones independientes se considera las medidas de calidad de diámetro y longitud de los palos de chupete, se realiza un test de Shapiro Wilks generalizada para observaciones multivariantes donde se obtiene un $p - valor < 0.05$ por lo cual se rechaza la hipótesis nula que esta muestra está normalmente distribuida, cabe recalcar que los gráficos de control estándar están diseñados bajo el supuesto de normalidad de las observaciones, por tal motivo los resultados arrojados por el gráfico T^2 podría estar sujetos al error Tipo I y II. Anomalías fueron reportadas al departamento de calidad en los días 35, 43, 49, 53 y 54 causadas por la ausencia del operador debido a enfermedad en los dos primeros días y por haber sido reemplazado por un menos experto en el tercero, sin embargo en los tres días un operador menos experto tomó esa posición, los dos últimos días corresponden a mala calidad del material de polopropileno usado para el formado de este plástico.

T² para observaciones independientes

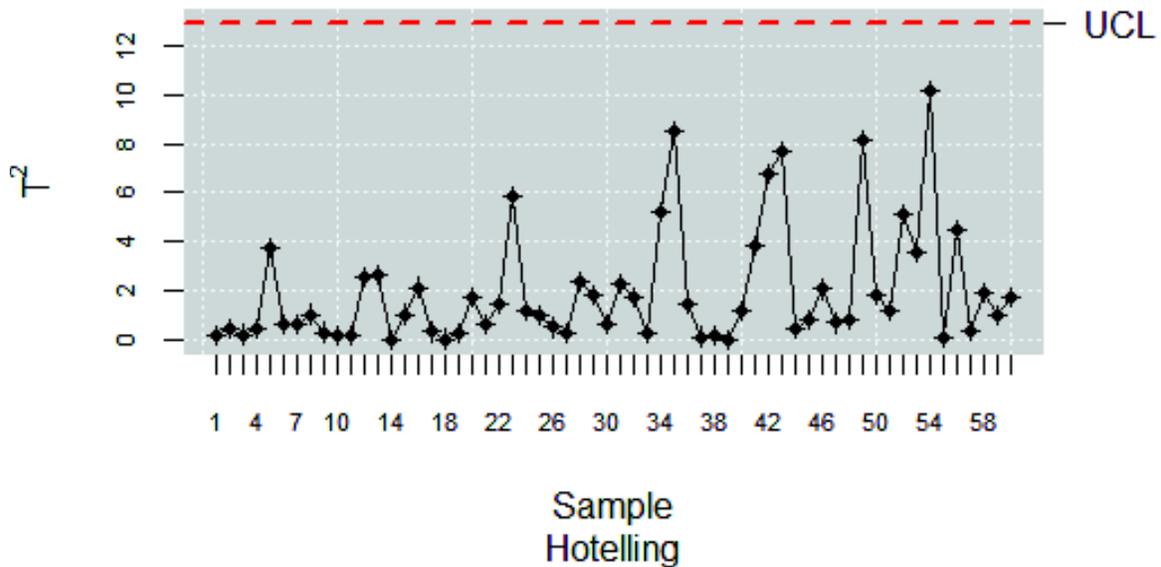


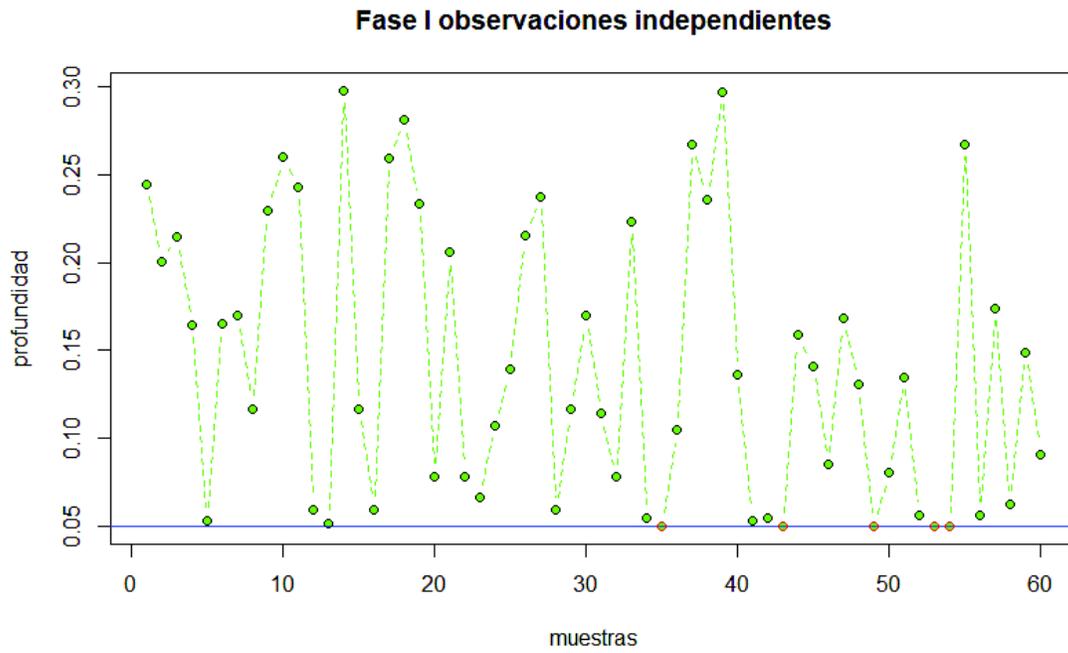
Figura 3.8: Gráfico T^2 de Hotelling para la Fase I, observaciones independientes

La figura 3.8 muestra el gráfico T^2 de Hotelling efectuado para la fase de calibrado de las observaciones, podemos notar que no detecta ningún punto fuera de control a pesar de que existen causas asignables reportadas.

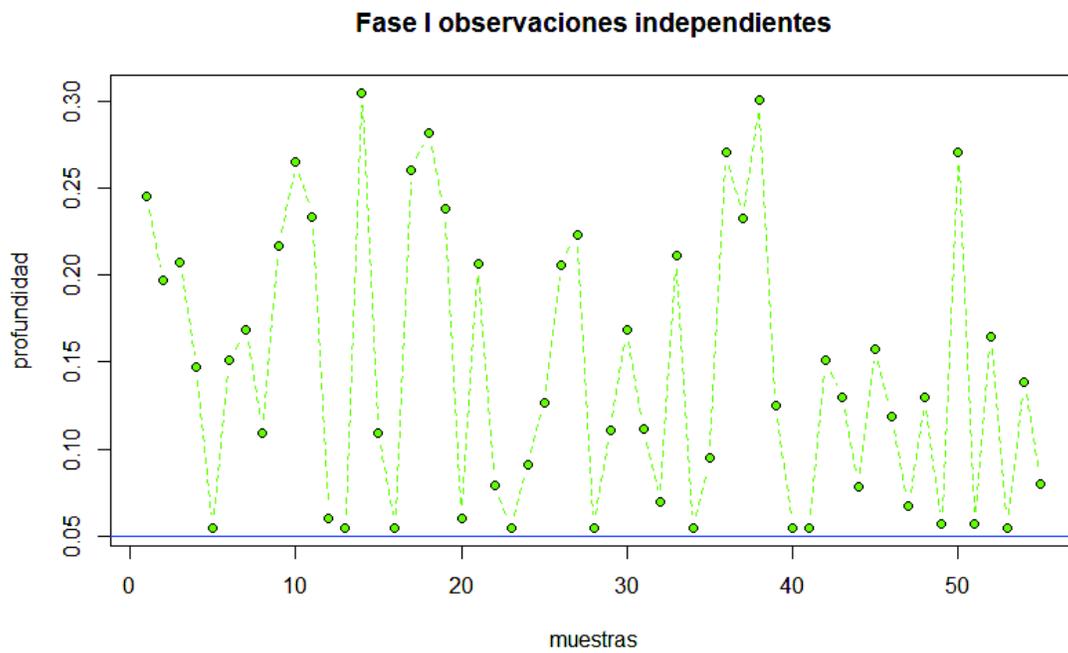
Por otro lado, en base al estudio de simulación presentado anteriormente para el gráfico de profundidades, se utiliza la profundidad simplicial y el método bootstrap suavizado estándar con coeficiente de suavizado $\gamma = 0.1$, además se fija un nivel de significancia $\alpha = 0.05$ para la estimación del LCI a partir de $B = 500$ muestras bootstrap.

En la parte superior de la figura 3.9 se muestra el gráfico de profundidad para cada día de producción, podemos notar como ahora este gráfico detecta las causas asignables previamente detectadas por lo que podemos empezar a evidenciar la capacidad de detección de los gráficos propuestos. A partir de esta muestra de calibrado se calcula el límite de control para la Fase I que servirá como determinante para el cálculo del porcentaje de rechazo para la Fase II, cabe recalcar que al descartar observaciones atípicas se parte de un proceso bajo control mostrado en la parte inferior de la figura 3.9 para la cual las causas asignables fueron removidas.

Una vez llevada a cabo la Fase I procedemos a la construcción de gráficos de control para la fase de monitoreo o Fase II. La muestra de monitoreo se compone por las



(a) Gráfico de control con observaciones atípicas



(b) Gráfico de control con observaciones en control, causas asignables fueron removidas

Figura 3.9: Gráfico de control para Fase I observaciones independientes

producciones del mes de Junio, estas nuevas observaciones son controladas comparando su profundidad respecto a la profundidad de las observaciones en control, el objetivo es medir la capacidad de detección del gráfico de profundidad comparado con el gráfico T^2 presentado en la figura 3.10.

T^2 para observaciones independientes, Fase II

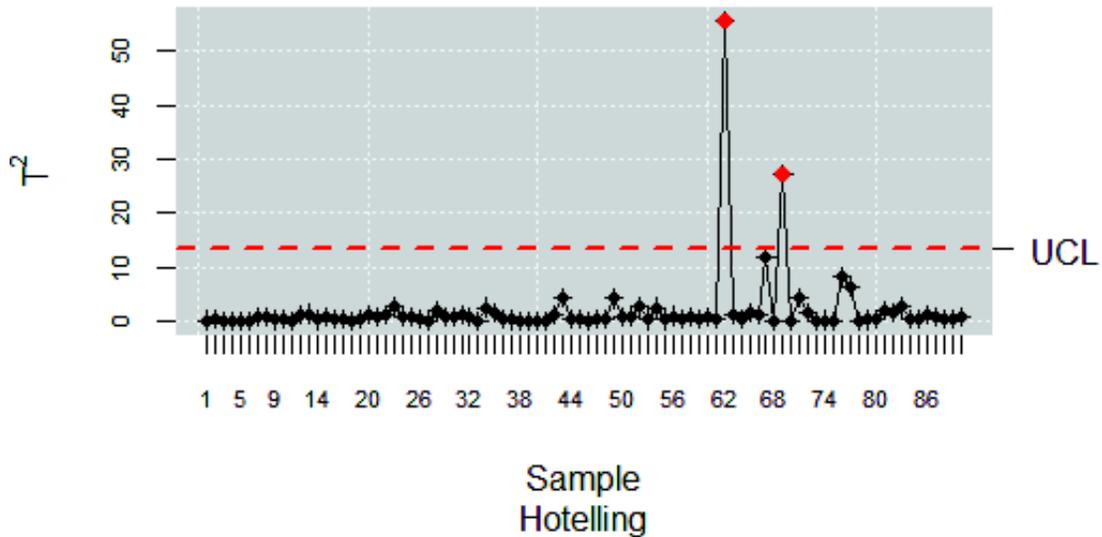


Figura 3.10: Gráfico T^2 de Hotelling para la Fase II, observaciones independientes

En la figura 3.11 se presenta el gráfico de profundidades para las muestras de referencia y monitoreo, como se puede notar este gráfico presenta una mejor capacidad de detección de atípicos. El mes de Junio representó varios problemas respecto al bajo peso de los chupetes, dado que la longitud de los palitos incide en el peso final de los mismos, unos se ocasionaron debido al caramelo, sin embargo otros se dieron por el plástico, se pudo identificar al menos 20 producciones en el mes donde se suscitaron estos inconvenientes, podemos notar como el gráfico T^2 de Hotelling solo identifica dos situaciones fuera de control, cuando nuestro gráfico propuesto identifica un gran porcentaje de ellas.

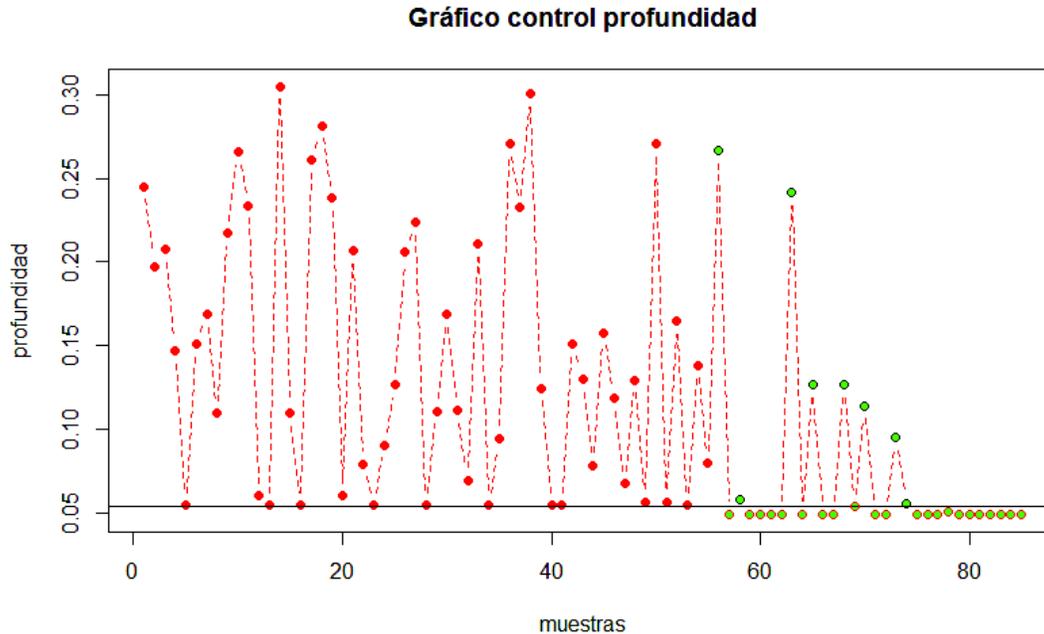


Figura 3.11: Gráfico de profundidad:muestra de referencia vs monitoreo para el caso independiente

Esto podemos verlo calculando la potencia del gráfico de control $1 - \beta$ para lo cual se obtiene para el caso de gráfico de control estándar un valor de 0.06, por otro lado para el gráfico propuesto un valor de 0.724 lo cual es un claro indicio de la mejoría al detectar atípicos que surgen en el proceso.

Caso Dependiente

El caso de estudio de las variables dependientes se basa en las medidas de calidad de peso y longitud de los palos de chupete, como se había mencionado anteriormente estas dos variables están correlacionadas dado que la longitud incide en el peso final, esto se verifica determinando el coeficiente de correlación entre ambas variables obteniéndose un valor de 0.5 lo que implica correlación positiva.

Realizando un test de Shapiro-Wilks generalizado obtenemos un $p - valor < 0.05$ por tanto rechazamos la hipótesis nula que los datos provienen de una distribución normal, así, no se cumple tanto la hipótesis de distribución normal para los gráficos de control estándar así como la independencia entre observaciones.

De manera similar al caso independiente se presentan anomalías reportadas por el departamento de calidad semejantes a las expuestas en el caso anterior suscitadas

por calidad en la materia prima y el personal humano encargado de la máquina.

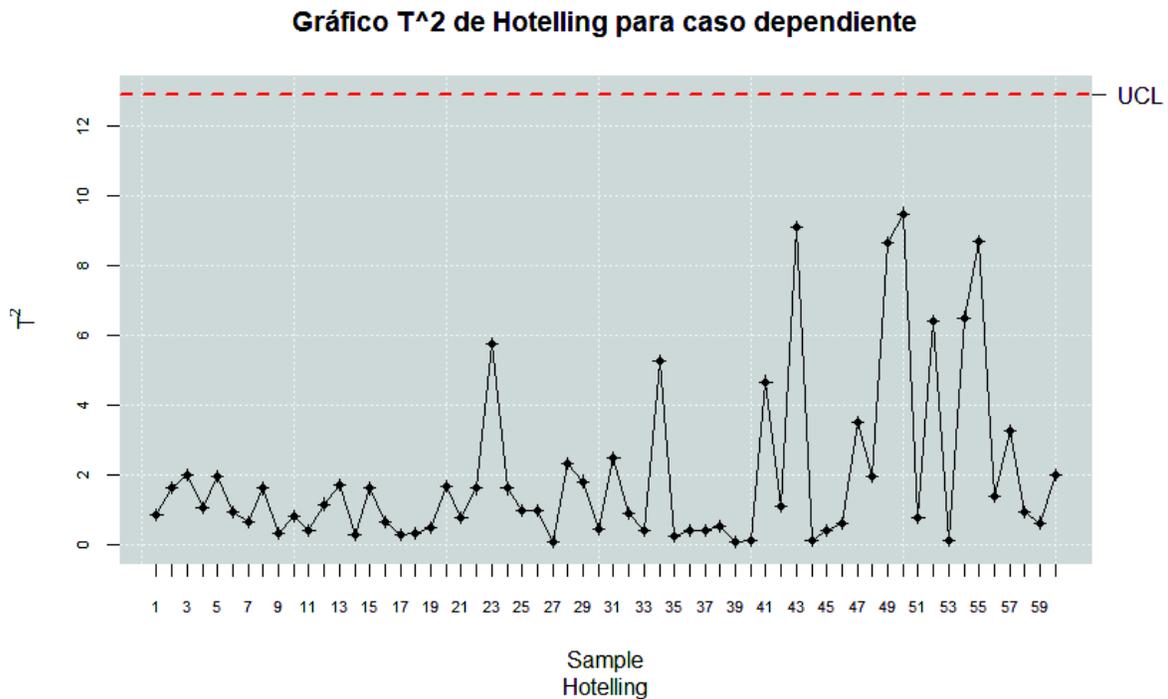
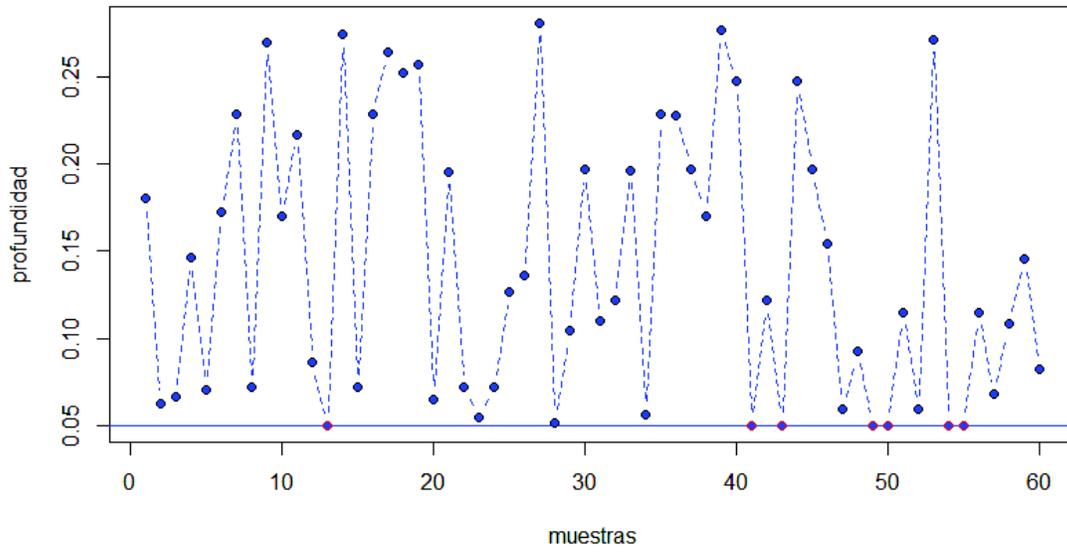


Figura 3.12: Gráfico T^2 de Hotelling para la Fase I, observaciones correlacionadas

La figura 3.12 muestra el gráfico el gráfico T^2 para el cual podemos notar que no detecta ningún punto fuera de control, esto era de esperarse al no cumplir suposiciones para el diseño de este gráfico, por otro lado para el diseño del gráfico de profundidades en base a los resultados expuestos en el estudio de simulación se usa la profundidad Simplicial justo con el método de bootstrap estacionario estándar con media de la distribución geométrica de los bloques igual a $b = 5$. Se fija un nivel de significancia $\alpha = 0.05$ para la estimación del LCI a partir de $B = 500$ muestras bootstrap.

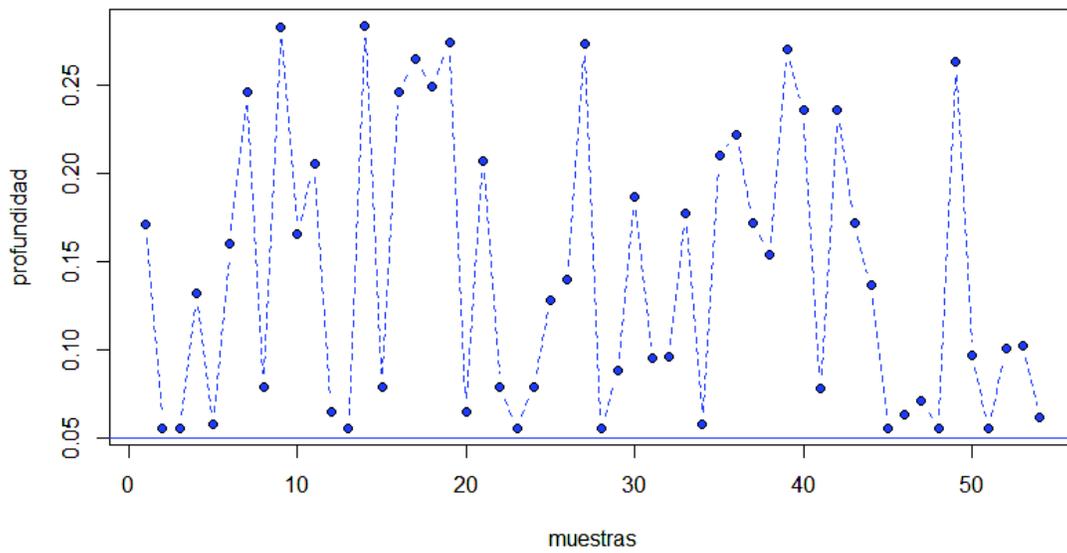
En la parte superior de la figura 3.13 apreciamos la detección de puntos atípicos y corroboramos la eficiencia en la detección así como en el caso independiente, una vez que se eliminan las causas asignables a partir de esta muestra de calibrado se calcula el límite de control para la Fase I que servirá como determinante para el cálculo del porcentaje de rechazo para la Fase II, cabe recalcar que al descartar observaciones atípicas se parte de un proceso bajo control mostrado en la parte inferior de la figura 3.13 para la cual las causas asignables fueron removidas.

Fase I observaciones correlacionadas



(a) Gráfico de control con observaciones atípicas

Fase I observaciones correlacionadas



(b) Gráfico de control con observaciones en control

Figura 3.13: Gráfico de control para Fase I observaciones dependientes

Partiendo de un proceso en control, se construyen los gráficos para la Fase II o de monitoreo, como se había comentado en el mes de Junio se suscitaron un sinnúmero de problemas referente a la calidad del producto terminado que no fueron detectados por los índices de calidad propuestos por el departamento para detectar desviaciones en la calidad, el gráfico 3.14 muestra la detección del gráfico T^2 el que solamente detecta dos puntos fuera de control.

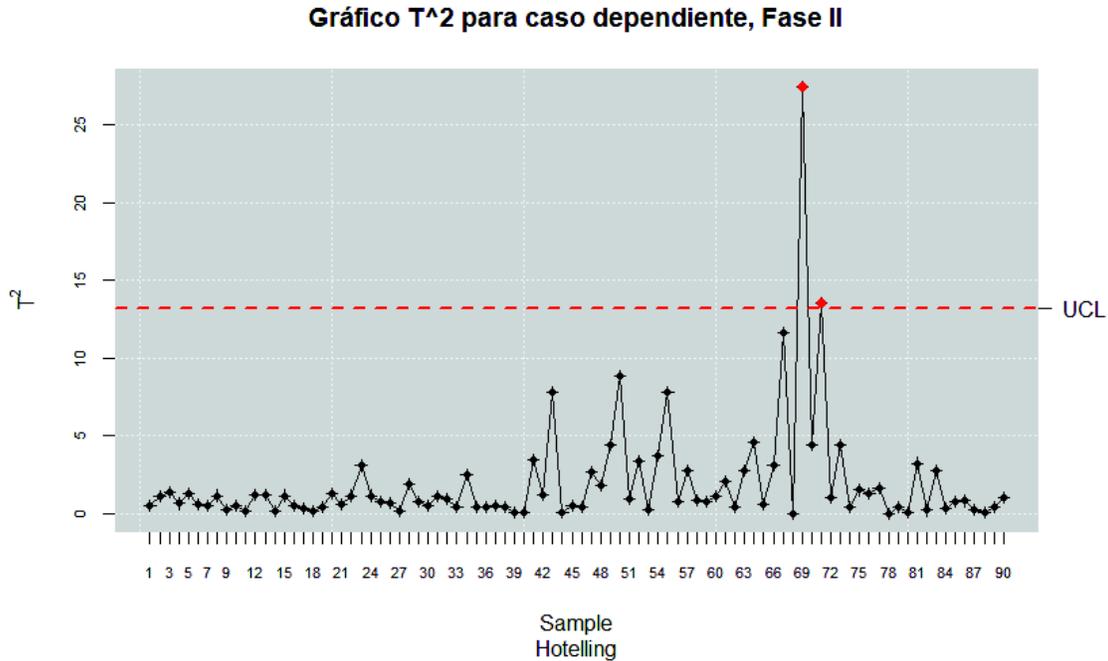


Figura 3.14: Gráfico T^2 de Hotelling para la Fase II, observaciones dependientes

En contraste, en la figura 3.15 apreciamos una mayor capacidad de detección de puntos atípicos, lo que implica una mejoría en el control de procesos, de igual manera es seguro que detecta cambios en la magnitud del proceso que los gráficos estándar no realizan.

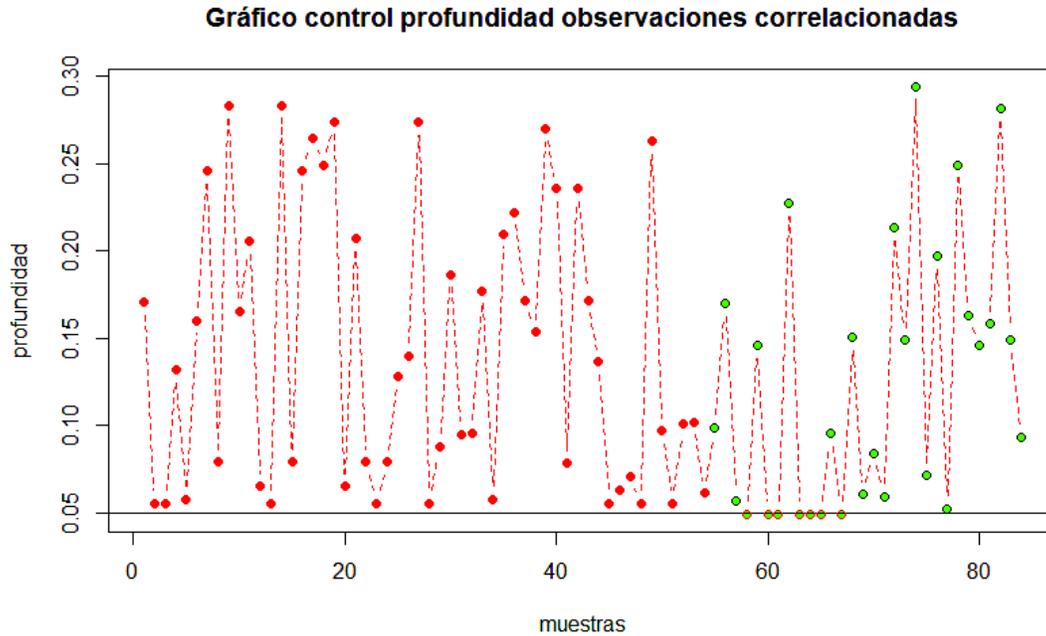


Figura 3.15: Gráfico de profundidad:muestra de referencia vs monitoreo para el caso dependiente

Como se ha podido observar en el estudio de simulación y en el caso de estudio, la nueva propuesta de gráficos de control para las Fases I y II es capaz de detectar de forma eficiente anomalías en una amplia gama de escenarios y bases de datos. Para el caso de la aplicación al caso de estudio de extruido de chupetes, la metodología empleada en este proyecto de investigación servirá como punto de partida para un proyecto de mejora continua de la calidad mediante monitoreo continuo a través del tiempo que permita controlar, monitorear y detectar anomalías en el proceso, cabe recalcar que esta alternativa se está introduciendo de a poco en los procesos de Confiteca C.A, más precisamente en la planta de plásticos donde varias medidas de calidad son determinantes para la calidad del producto final y que al usar metodologías estándar (las que se usa actualmente) no está rindiendo los frutos esperados y generando problemas a corto y mediano plazo.

Capítulo 4

Conclusiones y Recomendaciones

En el presente trabajo de investigación se ha propuesto una nueva alternativa de gráficos de control que permitan controlar la variabilidad de procesos o servicios y así detectar anomalías a partir de datos autocorrelacionados e independientes. La propuesta incluye alternativas para desarrollo en la Fase I para estabilizar el proceso y la Fase II para monitoreo de los gráficos de control, se desarrolla la Fase I a pesar de no estar incluido dentro de los objetivos al ver la necesidad de incluirla durante el proceso.

El desarrollo de ambas fases se basa en la detección de puntos atípicos y métodos bootstrap aplicados a gráficos de control no paramétricos basados en medidas de profundidad. La combinación de las dos fases dan lugar a la metodología propuesta, la cual fue aplicada con éxito a un caso de estudio de un proceso productivo. Además se realizó un estudio de simulación que permite comparar el desempeño medido por el porcentaje de rechazo cuando no se cumple la hipótesis nula de los gráficos dependiendo de la medida de profundidad utilizada.

Se comparó el uso de diferentes tipos de profundidad para el desarrollo de las Fases I y II contrastando contra los gráficos propuestos por Regina Liu mostrando la consistencia de los mismos, dado que al aumentar el tamaño de muestra aumenta la potencia de la prueba. Para la Fase I realizamos dos escenarios de simulación, uno para estimar y comparar el error Tipo I (tasa de falsas alarmas) y otro para evaluar la capacidad de detectar un cambio en la forma de la media del proceso a través de la estimación de su potencia. En el caso de datos independientes para el primer escenario de simulación para la Fase I notamos que para tamaños de muestra pequeños la medida de profundidad de Tukey arroja los mejores resultados acercándose al nivel de significancia fijado α , sin embargo conforme aumenta el tamaño de la muestra la

medida Simplicial arroja mejores resultados tanto para el caso de simulación Monte Carlo como para simulación bootstrap, el caso dependiente arroja resultados similares para todos los escenarios de coeficiente de correlación destacándose cuando se tiene un coeficiente ρ igual a 0.7 y 0.9.

Al evaluar el segundo escenario para la Fase I todas las medidas de profundidad a excepción de la de Tukey presentan buenos resultados (dada su lenta convergencia), sin embargo, la profundidad simplicial destaca al detectar pequeños cambios en la forma en el caso dependiente cabe destacar que para la mayoría de los casos mientras más pequeño sea el coeficiente de correlación mayor es la potencia de la prueba.

Respecto a la Fase II, se compara contra los gráficos Rank presentados por Liu, dado el acercamiento asintótico de este, presenta una menor potencia para tamaños de muestra pequeños comparado con los métodos propuestos, por otro lado para tamaños de muestra iguales a 50 y 100 la potencia de la prueba es semejante al del gráfico Rank lo cual es un indicio de la buena capacidad de detección de los gráficos propuestos, tanto para cuando se presenta cambios en la forma en una componente del vector de medias como en todo el vector.

En la práctica realizar el monitoreo mediante los gráficos de control propuestos representa un incremento en la capacidad de detección de atípicos con respecto a los gráficos convencionales lo cual presenta una ventaja para la eliminación de las principales causas asignables de la variabilidad del proceso, causante de errores y anomalías en el proceso.

Finalmente, se realizó un caso de estudio de control de calidad de extrusión de palitos de chupetes fabricados en Confiteca C.A para el cual se usa un índice de calidad para determinar la calidad global del producto, se mostró que tanto para las Fases I y II la capacidad de detección de atípicos es inmensamente mayor a los gráficos tradicionales (T^2 de Hotelling) dado que estos detectan pocos o ninguno de los atípicos reportados por el departamento de calidad que influyen en la calidad del producto terminado.

Para futuras líneas de investigación sería de interés la construcción de gráficos de control basados en los estadísticos Q (\bar{X}) y S (*CUSUM*) propuestos por Liu que permitan detectar cambios aún más pequeños en el proceso, además de incluir un estudio de simulación basado en cambio en forma y magnitud.

Por otro lado, sería valioso realizar estudios de simulación para el caso de datos autocorrelacionados usando bootstrap estacionario de la misma manera que en el

proyecto de investigación pero modificando la media de la distribución geométrica para generación de tamaño del bloque, más aún, sería relevante realizar un estudio de simulación mediante la técnica bootstrap de bloques móviles ajustando las observaciones a un modelo de series temporales para trabajar con un proceso estacionario y determinar la medida de m -dependencia y el tamaño de bloque.

Con respecto a futuras aplicaciones se aplicarán los gráficos de profundidad para monitorear el moldeo por inyección de plásticos usado para toda la categoría de caramelos interactivos de la marca Dr. Look.

Bibliografía

- Andrés M Alonso, Daniel Peña, and Juan Romo Urroz. Una revisión de los métodos de remuestreo en series temporales. 2002.
- FB Alt. Multivariate quality control: State of the art. In *ASQC Quality Congress Transactions—Detroit, MI*, pages 886–893, 1982.
- Layth C Alwan and Harry V Roberts. Time-series modeling for statistical process control. *Journal of Business & Economic Statistics*, 6(1):87–95, 1988.
- Daniel W Apley and Fugee Tsung. The autoregressive t 2 chart for monitoring univariate autocorrelated processes. *Journal of Quality Technology*, 34(1):80–96, 2002.
- Angel Francisco Arvelo. *Capacidad de Los Procesos Industriales: Métodos Estadísticos Exigidos Por Las Normas ISO-9000*. Universidad Católica Andrés, 1998.
- SM Bajgier. The use of bootstrap to construct limits on control charts. In *Proceedings of the decision Science Institute*, pages 1611–1613, 1992.
- Vic Barnett. The ordering of multivariate data. *Journal of the Royal Statistical Society. Series A (General)*, pages 318–355, 1976.
- F Breyfogle. Implementing six sigma part 1. In *The Quality Management Forum, ASQ, Summer*, pages 1–3, 1999.
- Ricardo Cao. Apuntes remuestreo. In *Técnicas de remuestreo*, pages 1–76, 2017.
- Edward Carlstein, Kim-Anh Do, Peter Hall, Tim Hesterberg, Hans R Künsch, et al. Matched-block bootstrap for dependent data. *Bernoulli*, 4(3):305–328, 1998.
- Ignacio Cascos, Angel López, and Juan Romo. Data depth in multivariate statistics. *Boletín de Estadística e Investigación Operativa*, 27(3):151–174, 2011.
- Emerson Chaparro and José Vargas. Gráficos de control para la media de un proceso en poblaciones con distribución asimétrica. *Revista Colombiana de Estadística*, 23(2): 29–44, 2000.

- Bradley Efron and Robert Tibshirani. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical science*, pages 54–75, 1986.
- Lucas Fernández. Algunas medidas de profundidad para datos multivariados y funcionales. *Universidad de Buenos Aires*, 2013.
- GestioPolis. ¿qué es seis sigma? metodología e implementación, 2001. URL <https://www.gestiopolis.com/que-es-seis-sigma-metodologia-e-implementacion/>.
- Peter Hall, Joel L Horowitz, and Bing-Yi Jing. On blocking rules for the bootstrap with dependent data. *Biometrika*, 82(3):561–574, 1995.
- Harold Hotelling. Multivariate quality control. *Techniques of statistical analysis*, 1947.
- W Jiang and K-L Tsui. An economic model for integrated apc and spc control charts. *Iie Transactions*, 32(6):505–513, 2000.
- LI Jing and BAO Changchun. A pitch detector based on the dyadic wavelet transform and the autocorrelation function. In *Signal Processing, 2002 6th International Conference on*, volume 1, pages 414–417. IEEE, 2002.
- L Allison Jones. The statistical design of ewma control charts with estimated parameters. *Journal of Quality Technology*, 34(3):277–288, 2002.
- Seoung Bum Kim, Weerawat Jitpitaklert, and Thuntee Sukchotrat. One-class classification-based control charts for monitoring autocorrelated multivariate processes. *Communications in Statistics—Simulation and Computation*®, 39(3):461–474, 2010.
- Hans R Kunsch. The jackknife and the bootstrap for general stationary observations. *The annals of Statistics*, pages 1217–1241, 1989.
- Daniel Lanhede. Non-parametric statistical process control: Evaluation and implementation of methods for statistical process control at ge healthcare, umeå, 2015.
- Regina Y Liu. On a notion of data depth based on random simplices. *The Annals of Statistics*, pages 405–414, 1990.
- Regina Y Liu. Control charts for multivariate processes. *Journal of the American Statistical Association*, 90(432):1380–1387, 1995.

- Regina Y Liu and Kesar Singh. A quality index based on data depth and multivariate rank tests. *Journal of the American Statistical Association*, 88(421):252–260, 1993.
- Regina Y Liu and Jen Tang. Control charts for dependent and independent measurements based on bootstrap methods. *Journal of the American Statistical Association*, 91(436):1694–1700, 1996.
- Cynthia A Lowry, William H Woodall, Charles W Champ, and Steven E Rigdon. A multivariate exponentially weighted moving average control chart. *Technometrics*, 34(1):46–53, 1992.
- Douglas C Montgomery. *Introduction to statistical quality control*. John Wiley & Sons, 2007.
- Douglas C Montgomery and Christina M Mastrangelo. Some statistical process control methods for autocorrelated data. *Journal of Quality Technology*, 23(3):179–193, 1991.
- Karl Mosler. Depth statistics. In *Robustness and complex data structures*, pages 17–34. Springer, 2013.
- ES Page. Cumulative sum charts. *Technometrics*, 3(1):1–9, 1961.
- Dimitris N Politis and Joseph P Romano. The stationary bootstrap. *Journal of the American Statistical association*, 89(428):1303–1313, 1994.
- S Psarakis and GEA Papaleonida. Spc procedures for monitoring autocorrelated processes. *Quality Technology & Quantitative Management*, 4(4):501–540, 2007.
- SW Roberts. Control chart tests based on geometric moving averages. *Technometrics*, 1(3):239–250, 1959.
- Germana Scepi. Parametric and non parametric multivariate quality control charts. In *Multivariate Total Quality Control*, pages 163–189. Springer, 2002.
- Germana Scepi and Antonio Acconcia. Non parametric control charts for sequential process. In *COMPSTAT*, pages 447–452. Springer, 1996.
- Walter Andrew Shewhart. *Economic control of quality of manufactured product*. ASQ Quality Press, 1931.
- Tim Stapenhurst. *Mastering statistical process control*. Routledge, 2013.

Don G Wardell, Herbert Moskowitz, and Robert D Plante. Run-length distributions of special-cause control charts for correlated processes. *Technometrics*, 36(1):3–17, 1994.

Sijbrand Jacobus Wierda. Multivariate statistical process control—recent results and directions for future research. *Statistica Neerlandica*, 48(2):147–168, 1994.

William H Woodall and Matoteng M Ncube. Multivariate cusum quality-control procedures. *Technometrics*, 27(3):285–292, 1985.

Yijun Zuo and Robert Serfling. General notions of statistical depth function. *Annals of statistics*, pages 461–482, 2000.

Anexos

A.1. Código en R para simulaciones propuestas por Liu

A.1.1. Simulación de gráficos de control para procesos multivariantes

```
rm(list = ls())
#install.packages("MASS")
library(MASS)
set.seed(12545)
#Generación de observación normal bivalente
sigma <- matrix(c(1,0,0,1),nrow = 2,ncol = 2,byrow = TRUE)
G <-mvrnorm(n = 540,mu = c(0,0),Sigma = sigma)
#Denotemos los primeros 500 como Yi, i=1,...,500
Y <-G[1:500,]
#Denotamos los últimos 40 como Xi i=1,...,40
X1<-G[501:540,]
#Generamos otra distribución normal bivalente
sigma2 <- matrix(c(4,0,0,4),nrow = 2,ncol = 2,byrow = TRUE)
X2 <-mvrnorm(n = 40,mu = c(2,2),Sigma = sigma2)
#Combinamos X1 y X2 en una observación X
X <-rbind(X1,X2)
#install.packages("fda.usc")
#Calculo de simplicial depth
library(fda.usc)
dep <-function(data,datam,indices){
  dep <- mdepth.SD(data[indices],datam)$dep
  return(dep)
```

```

}
#####
#Calculo de rank

rgm <-function(data,dat){
  rank<-numeric(nrow(data))
  dep<- mdepth.SD(dat,dat)$dep
  dep2<- mdepth.SD(data,dat)$dep
  for(i in 1:length(dep2))
  rank[i] <- sum(dep2[i]>=dep)/nrow(dat)
  return(rank)
}

rnk<-rgm(X,Y)
#Gráfico r
plot(rnk,type = "b",xlab="Observaciones",ylab="Rango")
abline(h=0.5,col="red")
abline(h=0.025,col="red")
text(70,0.55,"CL=0.5")
text(10,0.08,"LCL=0.025")
#####
#Gráfico Q
#Para n=4
n<-4
Q <- matrix(rnk,length(rnk)/n,n,byrow = TRUE)
qm <-apply(Q,1,mean)
plot(qm,type = "l",
      xlab="Observaciones",ylab="Promedio de Rango")
abline(h=0.5,col="red")
abline(h=0.22,col="red")
text(15,0.45,"CL=0.5")
text(5,0.08,"LCL=0.22")
#Para n=10
n<-10
Q <- matrix(rnk,length(rnk)/n,n,byrow = TRUE)
qm <-apply(Q,1,mean)

```

```

plot(qm,type = "l",
      xlab="Observaciones",ylab="Promedio de Rango")
#abline(h=0.5,col="red")
abline(h=0.3193,col="red")
text(2,0.28,"LCL=0.3193")
#text(7,0.45,"CL=0.5")
#####
#Gráfico S

s<-numeric(length(rnk))
LCL<-numeric(length(rnk))
for (i in 1:(length(rnk)-1)){
  s[1]<-rnk[1]-0.5
  s[i+1]<-s[i]+rnk[i+1]-0.5
  LCL[i+1]<- -1.96*sqrt(((i)^2*(1/500+1/i))/12)
}
plot(s,type = "l",
      xlab="Observaciones",ylab="Sumas acumuladas")
lines(LCL,col="red")

#####
#Gráfico S*

sa <-numeric(80)
for(i in 1:80){
  sa[i]<-s[i]/(sqrt(i^2*(1/500+1/i)/12))
}

plot(sa,type = "l",
      xlab="Observaciones",ylab="Sumas acumuladas estand.")
abline(h=-1.96,col="red")
text(60,-2.5,"LCL=-1.96")

```

A.1.2. Gráficos de control basados en métodos bootstrap

Primer ejemplo

```
rm(list = ls())
library(boot)
library(dplyr)
#Simulation Example A
set.seed(12123)
n<-100
x <-rnorm(n,0,1)
y <-rexp(n, rate = 1)

#Formamos subgrupos de cuatros observaciones consecutivas (k=25,n=4)
X <- matrix(x,25,4,byrow = TRUE)
Y <- matrix(y,25,4,byrow = TRUE)

#Calculamos el promedio de la muestra
xm <-mean(x)
ym <-mean(y)

#Calculamos el promedio y la desviación estándar de los subgrupos
sx <-apply(X,1,mean)
sy <-apply(Y,1,mean)
dx <-apply(X,1,sd)
dy <-apply(Y,1,sd)

#Obtenemos la muestra bootstrap
bx <-sample(x,size=length(x)*8,replace = TRUE) %>% matrix(nrow = 200)
by <-sample(y,size=length(x)*8,replace = TRUE) %>% matrix(nrow = 200)

#Cálculo de los límites de control
#Límites usuales
#Cálculo de  $\bar{X}$  y  $s^2$ 
xbb <-sum(sx)/nrow(X)
ybb <-sum(sy)/nrow(Y)
```

```

ddx <-sum(dx^2)/nrow(X)
ddy <-sum(dy^2)/nrow(Y)

#Confianza 90%,95%,98%
#Establecemos nivel de confianza
p <-qnorm(0.975)
LCLn <-xbb-p*(sqrt(ddx)/sqrt(ncol(X)))
UCLn <-xbb+p*(sqrt(ddx)/sqrt(ncol(X)))

eLCLn<-ybb-p*(sqrt(ddy)/sqrt(ncol(Y)))
eUCLn<-ybb+p*(sqrt(ddy)/sqrt(ncol(Y)))

#Límites Bootstrap
bbx<-sum(apply(bx,1,mean))/nrow(bx)
pr <-0.975
hjk <-sqrt(ncol(bx))*(xm-apply(bx,1,mean))
pb <-quantile(hjk,probs = pr)
pbb <-quantile(hjk,probs = 1-pr)

LCLb <-bbx+pbb/sqrt(ncol(bx))
UCLb <-bbx+pb/sqrt(ncol(bx))

bby<-sum(apply(by,1,mean))/nrow(by)
hjj <-sqrt(ncol(by))*(ym-apply(by,1,mean))
pb1 <-quantile(hjj,probs = pr)
pbb1 <-quantile(hjj,probs = 1-pr)

eLCLb <-bby+pbb1/sqrt(ncol(by))
eUCLb <-bby+pb1/sqrt(ncol(by))

#Gráfico para nivel de confianza 0,05
#Exponencial
#Límites de control exactos para nivel de confianza
eLCL <- 0.273
eUCL <- 2.192
plot(sy,type = "b",ylim = c(-1,3),ylab = "X-bar",xlab = "Time")

```

```

abline(h=eUCLn,col="blue")
text(25,2.1,"Std",col = "Blue")
abline(h=eLCLn,col="blue")
text(25,-0.05,"Std",col = "Blue")
abline(h=eUCLb,col="green")
text(25,1.8,"Btp",col = "Green")
abline(h=eLCLb,col="green")
text(25,-0.22,"Btp",col = "Green")
abline(h=eUCL,col="red")
text(25,2.3,"Exact",col = "Red")
abline(h=eLCL,col="red")
text(25,0.2,"Exact",col = "red")
#Normal
#Límites de control exactos para nivel de confianza
LCL <- -0.979
UCL <- 0.979
plot(sx,type = "b",ylab = "X-bar",xlab = "Time")
abline(h=UCLn,col="blue")
text(25,0.75,"Std",col = "Blue")
abline(h=LCLn,col="blue")
text(25,-0.9,"Std",col = "Blue")
abline(h=UCLb,col="green")
text(25,1.1,"Btp",col = "Green")
abline(h=LCLb,col="green")
text(25,-1.1,"Btp",col = "Green")
abline(h=UCL,col="red")
text(25,0.95,"Exact",col = "Red")
abline(h=LCL,col="red")
text(25,-1,"Exact",col = "red")

```

Segundo ejemplo

```

rm(list = ls())
set.seed(12125)
#Obtención de muestras de variables aleatorias
A <-rnorm(101)
y <-rnorm(n = 100,mean = 0,sd = 0.01)

```

```

x <-numeric(length(y))
for (i in 1:length(y))
  x[i]<-(A[i]+A[i+1])/2+y[i]

#Tamaño de bloque
#Bloque de tamaño 4
b <-4
#Bloques
B <-matrix(0,ncol = b,nrow = (length(x)-b+1))
for (i in 1:(length(x)-b+1)){
  B[i,]<-x[i:(i+b-1)]
}

#Moving Blocks Bootstrap
#Número de muestras bootstrap
N <-1000
#Matriz de bloques bootstrap
BB <-matrix(0,nrow=N,ncol=b)
BB<-B[sample(1:nrow(B),N,replace = TRUE),]

#Media observaciones Xi
mx <-mean(x)
#Medias matriz MBB
mbx <-apply(BB,1,mean)
#Cálculo de límites de control
alpha <-0.05
hbb <-sqrt(N)*(mbx-mx)
pbl <-quantile(hbb,probs = (alpha/2))
pbu <-quantile(hbb,probs = 1-(alpha/2))

LCL4 <-(sum(mbx)/length(mbx))+pbl/sqrt(N)
UCL4 <-(sum(mbx)/length(mbx))+pbu/sqrt(N)
#####
#Tamaño de bloque
#Bloque de tamaño 2

```

```

b <-2
#Bloques
B <-matrix(0,ncol = b,nrow = (length(x)-b+1))
for (i in 1:(length(x)-b+1)){
  B[i,]<-x[i:(i+b-1)]
}

#Moving Blocks Bootstrap
#Número de muestras bootstrap
N <-1000
#Matriz de bloques bootstrap
BB <-matrix(0,nrow=N,ncol=b)
BB<-B[sample(1:nrow(B),N,replace = TRUE),]

#Media observaciones Xi
mx <-mean(x)
#Medias matriz MBB
mbx <-apply(BB,1,mean)
#Cálculo de límites de control
alpha <-0.05
hbb <-sqrt(N)*(mbx-mx)
pbl <-quantile(hbb,probs = (alpha/2))
pbu <-quantile(hbb,probs = 1-(alpha/2))

LCL2 <-(sum(mbx)/length(mbx))+pbl/sqrt(N)
UCL2 <-(sum(mbx)/length(mbx))+pbu/sqrt(N)
#####
#Tamaño de bloque
#Bloque de tamaño 1
b <-1
#Bloques
B <-matrix(0,ncol = b,nrow = (length(x)-b+1))
for (i in 1:(length(x)-b+1)){
  B[i,]<-x[i:(i+b-1)]
}

```

```

}

#Moving Blocks Bootstrap
#Número de muestras bootstrap
N <-1000
#Matriz de bloques bootstrap
BB <-matrix(0,nrow=N,ncol=b)
BB<-B[sample(1:nrow(B),N,replace = TRUE),]

#Media observaciones Xi
mx <-mean(x)
#Cálculo de límites de control
alpha <-0.05
hbb <-sqrt(N)*(BB-mx)
pbl <-quantile(hbb,probs = (alpha/2))
pbu <-quantile(hbb,probs = 1-(alpha/2))

LCL1 <-(sum(mbx)/length(mbx))+pbl/sqrt(N)
UCL1 <-(sum(mbx)/length(mbx))+pbu/sqrt(N)

#Gráfico
#Obtenemos nuevas muestras
set.seed(32646)
A <-rnorm(101)
y <-rnorm(n = 100,mean = 0,sd = 0.01)
x <-numeric(length(y))
for (i in 1:length(y))
  x[i]<-(A[i]+A[i+1])/2+y[i]
X<-matrix(x,25,4,byrow = TRUE)

#Límites de control exactos
#la posición de los límites de control depende de si la covarianza
# es negativa o positiva
LCL <- -0.92193

```

```

UCL <- 0.92193
plot(apply(X,1,mean),type = "b",ylab = "X-bar",xlab = "Time",
      ylim = c(-1.4,1.7))
abline(h=UCL,col="blue")
text(25,0.9,"Std",col = "Blue")
abline(h=UCL4,col="red")
text(25,1.05,"\u2113=1",col = "red")
abline(h=UCL2,col="brown")
text(25,1.35,"\u2113=2",col = "brown")
abline(h=UCL1,col="green")
text(25,1.6,"\u2113=4",col = "green")
abline(h=LCL,col="blue")
text(25,-0.85,"Std",col = "Blue")
abline(h=LCL4,col="red")
text(25,-0.65,"\u2113=1",col = "red")
abline(h=LCL2,col="brown")
text(25,-1.01,"\u2113=2",col = "brown")
abline(h=LCL1,col="green")
text(25,-1.35,"\u2113=4",col = "green")

```

A.2. Código en R para simulación observaciones independientes y dependientes

Mediante variaciones en los parámetros de las funciones del código, se simula para distintas medidas de profundidad, tamaños de muestra y tipos de observaciones.

A.2.1. Funciones para cálculo de estadísticos

```

library(fda.usc)
library(boot)
library(MASS)
#####
#Función para el cálculo del estadístico rank, q, s y s* independiente y

```

```

dependiente
sim.md<-function( n = 100, p = 2, ro = 0.7, mu = NULL, dis=rnorm,
                 x = seq(0, 1, length = p),q=5,depend=NULL,
                 func = mdepth.MhD, data.ref = NULL){
  # Media
  if (is.null(mu)) mu <- rep(0,p)
  # Covarianzas
  x.dist <- as.matrix(dist(x))
  x.cov <- exp(-x.dist)
  # Factorización de la matriz de covarianzas
  U <- chol(x.cov)
  #set.seed(54321)
  z <- matrix(dis(n * p), nrow = p)
  if(depend){
    error <- t(U) %*% z
    y <- matrix(NA, nrow = p,ncol = n)
    y[,1] <- error[,1]
    for (i in 2:n) {
      y[,i] <- ro*error[,i-1]+error[,i]
    }
    y <- y + mu
  }
  else{
    error <- t(U) %*% z
    y<-error+mu
  }
  data.new <- t(y)
  if (is.null(data.ref)) data.ref <- data.new
  dep.ref <-func(data.ref)$dep
  dep <- func(data.new, data.ref)$dep
  rank <- sapply(dep,function(x) mean(dep.ref <= x))
  if(n %% q==0){
    q<-apply(matrix(dep,ncol=q,byrow=TRUE),1,mean)
    s<-n*(q-1/2)
    sas<-s/(sqrt(n/12))
  }
}

```

```

else{
  stop("Ingrese un valor de q divisible para n")
}

result <-list(data = data.new, data.ref = data.ref,
             x.cov = x.cov, z = z, dep = dep, rank = rank,
             q=q,s=s,sas=sas)

  return(result)
}
#####
#Función generadora Monte Carlo
sim.mc<-function(mc=1000,n=25,p=2,mu=NULL,ro=NULL,q=5,dataref=NULL
                ,func = mdepth.MhD,depend=NULL){
  sim<-list(mc)
  for (i in 1:mc){
    set.seed(i)
    sim[[i]]<-sim.md(n,p=p,mu=mu,ro=ro,func=func,depend=depend,
                    q=q,data.ref = dataref)
  }
  return(sim)
}
#####
#Función generador de LCI y porcentaje de rechazo
Lim<-function(data,alpha=0.05,LCI=NULL){
  if (is.null(LCI)) LCI<-quantile(data,probs = alpha)
  rp <-mean(data<=LCI)
  result <-list(LCI=LCI,RP=rp)
}
#####
# Función generadora de estadísticos bajo hipótesis nula o alternativa
#(rank,q,s,s*)
# bajo dependencia o independencia
# Se incluye cálculo de porcentaje de rechazo y tabla de resumen
sim.esc<-function(ni,m,mc=1000,alpha=0.05,p=2,qi=5,mu=NULL,ro=NULL,
                 dataref=NULL,

```

```

        depend=NULL,
        func=mdepth.MhD,statistic){
if(depend){
  table<-data.frame(mu = rep(delta, each=M*K), ro = rep(rep(ro, each=M),U),
                    m = rep(rep(m,M),U),n = rep(rep(n[which(n==ni)],M*K),U),
                    LCI= numeric(M*K*U),LCImc = numeric(M*K*U),
                    LCIboot = numeric(M*K*U),
                    PRmc=numeric(M*K*U),PRboot=numeric(M*K*U))

c<-1
list.dep<-list(M*K*U)
list.rank<-list(M*K*U)
list.data<-list(M*K*U)
list.q<-list(M*K*U)
list.s<-list(M*K*U)
list.sas<-list(M*K*U)
for (k in 1:U){
  for (j in 1:K){
    for (i in 1:M){
      simmc<-sim.mc(mc=mc,n=m[i],p=p,mu=mu[,k],ro=ro[j],depend=depend,q=qi,
                   func=func,dataref=dataref[,mc])
      list.data[[c]] <- array(sapply(simmc,function(x) x$data),dim=c(m[i],p,mc))
      switch(statistic,
             dep={list.dep[[c]] <- matrix(sapply(simmc,function(x) x$dep),
                                           nrow=mc,
                                           ncol=m[i],byrow=TRUE)
             list.rank[[c]] <- matrix(sapply(simmc,function(x) x$rank),nrow=mc,
                                       ncol=m[i],byrow=FALSE)

      table[c,5]<-alpha
      table[c,6]<-LCI
      table[c,8]<-Lim(list.dep[[c]],LCI)$RP
      c<-c+1},
      rank={list.rank[[c]] <- matrix(sapply(simmc,function(x) x$rank),
                                       nrow=mc,ncol=m[i],byrow=TRUE)

      table[c,5]<-alpha
      table[c,6]<-LCI
      table[c,8]<-Lim(list.rank[[c]],LCI)$RP

```

```

c<-c+1},
q={list.q[[c]]<- matrix(sapply(simmc,function(x) x$q),nrow=mc,
                        ncol=(m[i]/qi),byrow=TRUE)

if(q<=5){
  LCIt <- ((factorial(qi)*alpha)^(1/qi))/qi
} else{
  LCIt<-0.5-qnorm(1-alpha)*(1/sqrt(12*qi))
}
table[c,5]<-LCIt
table[c,6]<-LCI
table[c,8]<-Lim(list.q[[c]],LCI)$RP
c<-c+1},
s={list.s[[c]]<- matrix(sapply(simmc,function(x) x$s),nrow=mc,
                        ncol=(m[i]/qi),byrow=TRUE)
LCIt<- -(qnorm(1-alpha)*(sqrt(qi/12)))
table[c,5]<- LCIt
table[c,6]<-LCI
table[c,8]<-Lim(list.s[[c]],LCI)$RP
c<-c+1},
sas={list.sas[[c]]<- matrix(sapply(simmc,function(x) x$sas),
                            nrow=mc,
                            ncol=(m[i]/qi),byrow=TRUE)
table[c,5]<- -qnorm(1-alpha)
table[c,6]<-LCI[which(n==ni)]
table[c,8]<-Lim(list.sas[[c]],LCI)$RP
c<-c+1},
stop("Ingrese estadístico apropiado")
)
}
}
}
}
else{
  table<-data.frame(mu = rep(delta, each=M),m = rep(rep(m,M),M),
                   n = rep(rep(n[which(n==ni)],M),U),LCI= numeric(M*U),
                   LCImc =

```

```

        numeric(M*U),
        LCIboot = numeric(M*U),PRmc=numeric(M*U),
        PRboot=numeric(M*U))
c<-1
list.dep<-list(M*U)
list.rank<-list(M*U)
list.data<-list(M*U)
list.q<-list(M*U)
list.s<-list(M*U)
list.sas<-list(M*U)
for (k in 1:U){
  for (i in 1:M){
    simmc<-sim.mc(mc=mc,n=m[i],p=p,mu=mu[,k],depend=depend,q=qi,
                 func=func,dataref=dataref[, ,mc])
    list.data[[c]] <- array(sapply(simmc,function(x) x$data),
                          dim=c(m[i],p,mc))
    switch(Statistic,
          dep={list.dep[[c]] <- matrix(sapply(simmc,function(x) x$dep),
                                       nrow=mc,
                                       ncol=m[i],byrow=TRUE)
          list.rank[[c]] <- matrix(sapply(simmc,function(x) x$rank),
                                   nrow=mc,
                                   ncol=m[i],byrow=TRUE)
          table[c,4]<-alpha
          table[c,5]<-LCI
          table[c,7]<-Lim(list.dep[[c]],LCI)$RP
          c<-c+1},
          rank={list.rank[[c]] <- matrix(sapply(simmc,
                                                  function(x) x$rank),
                                           nrow=mc,ncol=m[i],byrow=TRUE)
          table[c,4]<-alpha
          table[c,5]<-LCI
          table[c,7]<-Lim(list.rank[[c]],LCI)$RP
          c<-c+1},
          q={list.q[[c]]<- matrix(sapply(simmc,
                                          function(x) x$q),nrow=mc,

```

```

                                ncol=(m[i]/qi),byrow=TRUE)
if(q<=5){
  LCIt <- ((factorial(qi)*alpha)^(1/qi))/qi
} else{
  LCIt<-0.5-qnorm(1-alpha)*(1/sqrt(12*qi))
}
table[c,4]<-LCIt
table[c,5]<-LCI
table[c,7]<-Lim(list.q[[c]],LCI)$RP
c<-c+1},
s={list.s[[c]]<- matrix(sapply(simmc,
                                function(x) x$s),nrow=mc,
                                ncol=(m[i]/qi),byrow=TRUE)
LCIt<- -(qnorm(1-alpha)*(sqrt(qi/12)))
table[c,4]<- LCIt
table[c,5]<-LCI
table[c,7]<-Lim(list.s[[c]],LCI)$RP
c<-c+1},
sas={list.sas[[c]]<- matrix(sapply(simmc,
                                function(x) x$sas),
                                nrow=mc,ncol=(m[i]/qi),
                                byrow=TRUE)
table[c,4]<- -qnorm(1-alpha)
table[c,5]<-LCI
table[c,7]<-Lim(list.sas[[c]],LCI)$RP
c<-c+1},
stop("Ingrese estadístico apropiado")
)
}
}
}
result<-list(data=list.data,rank=list.rank,dep=list.dep,
             q=list.q,s=list.s,sas=list.sas,table=table)
return(result)
}
#####

```

```

# Función para cálculo de estadísticos bajo hipótesis nula
# y bajo datos de referencia
tsrank<-function(data,ind,data.ref=NULL,func=mdepth.MhD,q=5,depend=NULL,stat){
  nc<-nrow(data)
  if(depend){
    if(!is.null(data.ref)){
      dep.ref<-func(data.ref)$dep
      dep <- func(data, data.ref)$dep
      rank <- sapply(dep,function(x) mean(dep.ref <= x))
      if(nc %% q==0){
        q<-apply(matrix(dep,ncol=q,byrow=TRUE),1,mean)
        s<-nc*(q-1/2)
        sas<-s/(sqrt(nc/12))
      }
    }
    else{
      stop("Ingrese un valor de q divisible para n")
    }
    switch(stat,
           rank={return(rank)
                },
           q={return(q)
              },
           s={return(s)
              },
           sas={return(sas)
                },
           stop("Ingrese estadístico válido")
          )
  }
  else{
    dep <- func(data)$dep
    rank <- sapply(dep,function(x) mean(dep <= x))
    if(nc %% q==0){
      q<-apply(matrix(dep,ncol=q,byrow=TRUE),1,mean)
      s<-nc*(q-1/2)
      sas<-s/(sqrt(nc/12))
    }
  }
}

```

```

}
else{
  stop("Ingrese un valor de q divisible para n")
}
switch(stat,
  rank={return(rank)
  },
  q={return(q)
  },
  s={return(s)
  },
  sas={return(sas)
  },
  stop("Ingrese estadístico válido")
)
}
}
else{
  if(!is.null(data.ref)){
    dep.ref<-func(data.ref)$dep
    dep <- func(data[ind,], data.ref)$dep
    rank <- sapply(dep,function(x) mean(dep.ref <= x))
    if(nc %% q==0){
      q<-apply(matrix(dep,ncol=q,byrow=TRUE),1,mean)
      s<-nc*(q-1/2)
      sas<-s/(sqrt(nc/12))
    }
    else{
      stop("Ingrese un valor de q divisible para n")
    }
  }
  switch(stat,
    rank={return(rank)
    },
    q={return(q)
    },
    s={return(s)

```

```

        },
        sas={return(sas)}
    },
    stop("Ingreso estadístico válido")
)
}
else{
    dep <- func(data[ind,])$dep
    rank <- sapply(dep,function(x) mean(dep <= x))
    if(nc %% q==0){
        q<-apply(matrix(dep,ncol=q,byrow=TRUE),1,mean)
        s<-nc*(q-1/2)
        sas<-s/(sqrt(nc/12))
    }
    else{
        stop("Ingrese un valor de q divisible para n")
    }
    switch(stat,
        rank={return(rank)}
        },
        q={return(q)}
        },
        s={return(s)}
        },
        sas={return(sas)}
        },
        stop("Ingreso estadístico válido")
    )
}
}
}
#####
# Función generadora muestras bootstrap para estadísticos dependientes
# (bootstrap estacionario) e independientes
bootmc<-function(ni=25,n=n,data,mc=1000,nboot=500,
    data.ref=NULL,statistic=tsrank,

```

```

        q=5,depend=NULL,stat=NULL){
if(depend){
  bootp<-list(mc)
  for (i in 1:mc){
    rnkboot<-tsboot(tseries =data[,i], statistic = statistic,
                    R = nboot,sim = "geom",depend=TRUE,stat=stat,q=q,
                    data.ref=data.ref[, ,mc])$t
    bootp[[i]]<-rnkboot
  }
  if(stat=="dep"){
    matboot<-matrix(NA,nrow=(nboot*mc),ncol=ni)
    matboot<-matrix(unlist(bootp),ncol=ni)
    return(matboot)
  }
  else{
    matboot<-matrix(NA,nrow=(nboot*mc),ncol=(ni/q))
    matboot<-matrix(unlist(bootp),ncol=(ni/q))
    return(matboot)
  }
}
else{
  bootp<-list(mc)
  for (i in 1:mc){
    rnkboot<-boot(data=data[,i], statistic = statistic,q=q,
                  R = nboot,depend=FALSE,stat=stat,
                  data.ref=data.ref[, ,mc])$t
    bootp[[i]]<-rnkboot
  }
  if(stat=="dep"){
    matboot<-matrix(NA,nrow=(nboot*mc),ncol=ni)
    matboot<-matrix(unlist(bootp),ncol=ni)
    return(matboot)
  }
  else{
    matboot<-matrix(NA,nrow=(nboot*mc),ncol=(ni/q))
    matboot<-matrix(unlist(bootp),ncol=(ni/q))

```

```

        return(matboot)
    }
}
}
bootmc2<-function(ni=25,n=n,indi=1,data,mc=1000,nboot=500,
                  data.ref=NULL,statistic=tsrank,
                  q=5,depend=NULL,stat=NULL){
if(depend){
  bootp<-list(mc)
  for (i in 1:mc){
    rnkboot<-tsboot(tseries =data[[indi]][,i], statistic = statistic,
                    R = nboot,sim = "geom",depend=TRUE,stat=stat,q=q,
                    data.ref=data.ref[, ,mc])$t
    bootp[[i]]<-rnkboot
  }
  if(stat=="dep"){
    matboot<-matrix(NA,nrow=(nboot*mc),ncol=ni)
    matboot<-matrix(unlist(bootp),ncol=ni)
    return(matboot)
  }
  else{
    matboot<-matrix(NA,nrow=(nboot*mc),ncol=(ni/q))
    matboot<-matrix(unlist(bootp),ncol=(ni/q))
    return(matboot)
  }
}
else{
  bootp<-list(mc)
  for (i in 1:mc){
    rnkboot<-boot(data=data[[indi]][,i], statistic = statistic,q=q,
                  R = nboot,depend=FALSE,stat=stat,
                  data.ref=data.ref[, ,mc])$t
    bootp[[i]]<-rnkboot
  }
  if(stat=="dep"){
    matboot<-matrix(NA,nrow=(nboot*mc),ncol=ni)

```

```

    matboot<-matrix(unlist(bootp),ncol=ni)
    return(matboot)
  }
  else{
    matboot<-matrix(NA,nrow=(nboot*mc),ncol=(ni/q))
    matboot<-matrix(unlist(bootp),ncol=(ni/q))
    return(matboot)
  }
}
}
#####
# Simulación bootstrap dependiente o independiente dada data de referencia
# bajo hipótesis nula o alternativa de escenarios
bootalt<-function(list,mc=1000,nboot=500,ni,indi=1,m=m,
                  data.ref=NULL,statistic=tsrank,q=5,depend=NULL,stat){
  if(depend){
    bootalt<-list(U*K*M)
    d<-1
    pb = txtProgressBar(min = 0, max = 81, style = 3)
    for (j in 1:(U*K)){
      setTxtProgressBar(pb, d - 0.5)
      for (i in 1:M) {
        bootalt[[d]]<-bootmc2(ni=m[i],n=m,indi=d,data=list,mc=mc,nboot=nboot,
                              depend=depend,
                              stat=stat,data.ref = dataref)
        d<-d+1
      }
      setTxtProgressBar(pb, d)
    }
    close(pb)
    return(bootalt)
  }
  else{
    bootalt<-list(U*M)
    d<-1
    pb = txtProgressBar(min = 0, max = 27, style = 3)

```

```

for (j in 1:U){
  setTxtProgressBar(pb, d - 0.5)
  for (i in 1:M) {
    bootalt[[d]]<-bootmc2(ni=m[i],n=m,indi=d,data=list,mc=mc,nboot=nboot,
                        depend=depend,
                        stat=stat,data.ref = dataref)

    d<-d+1
  }
  setTxtProgressBar(pb, d)
}
close(pb)
return(bootalt)
}
}
#####
# Potencia de la prueba
PP<-function(list,LCI,depend=NULL){
  if(depend){
    rp<-numeric(U*M*K)
    d<-1
    for (j in 1:U){
      for (k in 1:K){
        for (i in 1:M){
          rp[d]<-Lim(list[[d]],LCI=as.numeric(LCI))$RP
          d<-d+1
        }
      }
    }
    return(rp)
  }
  else{
    rp<-numeric(U*M)
    d<-1
    for (j in 1:U){
      for (i in 1:M){
        rp[d]<-Lim(list[[d]],LCI=as.numeric(LCI))$RP

```

```

        d<-d+1
    }
}
return(rp)
}

}
PPnew<-function(list,LCIt){
  rp<-numeric(U*M)
  d<-1
  for (j in 1:U){
    for (i in 1:M){
      rp[d]<-Lim(list[[d]],LCI=as.numeric(LCIt))$RP
      d<-d+1
    }
  }
  return(rp)
}

#####
# Comparación porcentaje de rechazo Monte Carlo y bootstrap
comp<-function(table,ni,alpha=0.05,depend,delta){
  if(depend){
    for (k in 1:M){
      result.a<-table[table[,3]==m[k],]
      for (i in 1:K){
        filter <- result.a[result.a[,2] == ro[i],]
        y<-seq(0,max(filter[,7],filter[,8],filter[,9])+0.01,
              length.out = length(delta))
        plot(x = delta, y =y,
             type = "n",col = "brown", main = paste("ro =",ro[i],"m =", m[k],
             "n =", n[which(n==ni)]),
             xlab="",ylab="")
        legend(x=1,y=0.05,
              legend = c("PRliu","PRmc","PRboot"),
              col =c("blue","red","green"),lty = 5,bty = "n",cex=1)
        lines(filter[,7], col ="blue",x = delta)

```

```

        lines(filter[,8], col ="blue")
        lines(filter[,9], col ="red")
        abline(h = alpha, lty = 3)
    }
}
par(mfrow=c(1,1))
}
else{
  for (k in 1:M){
    filter<-table[table[,2]==m[k],]
    y<-seq(0,max(filter[,7],filter[,8],filter[,9])+0.01,
           length.out = length(delta))
    plot(x = delta, y =y,
         type = "n",col = "brown", main = paste("m =", m[k],
                                                "n =", n[which(n==ni)]))
        ,xlab="",ylab="")
    legend(x=1,y=0.4,
           legend = c("PRliu","PRmc","PRboot"),
           col =c("blue","red","green"),lty = 5,bty = "n",cex=1)
    lines(filter[,7], col ="blue",x = delta)
    lines(filter[,8], col ="red",x = delta)
    lines(filter[,9], col ="green",x = delta)
    abline(h = alpha, lty = 3)
  }
  par(mfrow=c(1,1))
}
}
#####
# Tabla para comparación de límite de control inferior asintótico,
Monte Carlo y bootstrap
tablalpha<-function(data,databoot,alph,stat,qi){
  A<-length(alph)
  switch (stat,
         rank = {Alpha<-alph},
         q={if(q<=5){
           Alpha <- ((factorial(qi)*alph)^(1/qi))/qi

```

```

    } else{
      Alpha <-0.5-qnorm(1-alpha)*(1/sqrt(12*qi))
    }
  },
  s={Alpha<- -(qnorm(1-alpha)*(sqrt(qi/12)))},
  sas={Alpha<- -qnorm(1-alpha)},
  stop("Ingrese estadístico válido")
)

tablalpha <- data.frame(n=rep(n,each=A),alpha=rep(Alpha,N),
                        LCImc=numeric(A*N),LCIboot=numeric(A*N))

c<-1
for (i in 1:N){
  for (j in 1:A){
    #montecarlo
    tablalpha[c,3]<-Lim(data[[i]],alpha=Alpha[j])$LCI
    #bootstrap
    tablalpha[c,4]<-Lim(databoot[[i]],alpha=Alpha[j])$LCI
    c<-c+1
  }
}
result<-list(table=tablalpha,alpha=alph)
return(result)
}

#####
# Bootstrap suavizado para datos independientes
bootker<-function(data,mc,nboot,h=0.00001,data.ref=NULL,
                  func=mdepth.MhD){
  n<-nrow(data)
  rs<-matrix(NA,ncol=n,nrow=nboot)
  rss<-numeric(n)
  rsu<-matrix(NA,ncol=n,nrow=(nboot*mc))
  boot<-list(mc)
  for (i in 1:mc){
    for (k in 1:nboot)
    {

```

```

    if(!is.null(data.ref)){
      dep<-func(data[, ,i],data.ref[, ,mc])$dep
    }
    else{
      dep<-func(data[, ,i])$dep
    }
    rss<-sample(dep,n,replace=T)
    rs[k,]<-rss+h*rnorm(n,0,1)

  }
  boot[[i]]=rs
}
rsu<-matrix(unlist(boot),ncol=n)
return(rsu)
}
bootker2<-function(data,ind=1,mc,nboot,h=0.00001,data.ref=NULL,
                    func=mdepth.MhD){
  n<-nrow(data[[ind]])
  rs<-matrix(NA,ncol=n,nrow=nboot)
  rss<-numeric(n)
  rsu<-matrix(NA,ncol=n,nrow=(nboot*mc))
  boot<-list(mc)
  for (i in 1:mc){
    for (k in 1:nboot)
    {
      if(!is.null(data.ref)){
        dep<-func(data[[ind]][, ,i],data.ref[, ,mc])$dep
      }
      else{
        dep<-func(data[[ind]][, ,i])$dep
      }
      rss<-sample(dep,n,replace=T)
      rs[k,]<-rss+h*rnorm(n,0,1)

    }
    boot[[i]]=rs
  }
}

```

```

}
rsu<-matrix(unlist(boot),ncol=n)
return(rsu)
}
bootaltker<-function(data,mc,ind=1,nboot,h=0.00001,data.ref=NULL,
                    func=mdepth.MhD){
  bootalt<-list(U*M)
  d<-1
  pb = txtProgressBar(min = 0, max = 27, style = 3)
  for (j in 1:U){
    setTxtProgressBar(pb, d - 0.5)
    for (i in 1:M) {
      bootalt[[d]]<-bootker2(data,ind=d,mc,nboot,h=0.00001,data.ref=NULL,
                            func=mdepth.MhD)
      d<-d+1
    }
    setTxtProgressBar(pb, d)
  }
  close(pb)
  return(bootalt)
}
#####
#### Variación aislada a observaciones
vais<-function(data,n,mu=2,mc=1000,func=func){
  for (i in 1:mc){
    set.seed(i)
    data[, , i][n,]<-sim.md(n=1,q=1,mu=c(mu,0),depend=FALSE)$data
    dep[i,]<-func(data[, , i])$dep
  }
  result<-list(data=data,dep=dep)
  return(result)
}
#####
### Eliminar porcentaje de observaciones de una muestra
trdat<-function(data,n,per=0.9,mc=1000,p=2,func=func){
  xdat<-array(NA,dim=c(n,p,mc))

```

```

dep<-matrix(NA,nrow=mc,ncol=n,byrow=TRUE)
rank<-matrix(NA,nrow=mc,ncol=n,byrow=TRUE)
for (i in 1:mc){
  x1<-data[, ,i][sample(1:nrow(data[, ,i]),nrow(data[, ,i])*per),]
  x2<-rbind(x1,x1[sample(1:nrow(x1),ceiling(nrow(data[, ,i])*(1-per))
            ,replace = FALSE),])
  xdat[, ,i]<-x2
  dep[i,]<-func(xdat[, ,i])$dep
  rank[i,] <- sapply(dep[i,],function(x) mean(dep[i,] <= x))
}
result<-list(data=xdat,dep=dep,rank=rank)
return(result)
}

#####
#Densidad profundidades

dendep<-function(res.sim,alpha=0.05,n,t){
  N<-length(n)
  for (i in 1:N){
    sdep<-res.sim[[i]][[t[1]]]; lc.prmc<-Lim(sdep,alpha = alpha)
    srnk<-res.sim[[i]][[t[2]]]
    bootdep<-res.sim[[i]][[t[3]]]; lc.prboot<-Lim(bootdep,alpha = alpha)
    ### Densidad
    leg<-plot(density(sdep),main = paste("n= ",n[i],"alpha= ",alpha),
             xlab="profundidad",
             ylim=c(0,2),col=2)
    legend("topleft",legend = c(paste("MC"),paste("Bootstrap")), bty = "n",
          col = c(2,3),lty = 1)
    lines(density(bootdep),main = paste("n= ",n[i],"alpha= ",alpha),
         xlab="profundidad",
         ylim=c(0,2),col=3)
  }
}

#####
##### Grafico de control de calidad

```

```

ccdep<-function(res.sim,alpha=0.05,n,t){
  N<-length(n)
  for (i in 1:N){
    sdep<-res.sim[[i]][[t[1]]]; lc.prmc<-Lim(sdep,alpha = alpha)
    srnk<-res.sim[[i]][[t[2]]]
    bootdep<-res.sim[[i]][[t[3]]]; lc.prboot<-Lim(bootdep,alpha = alpha)
    plot(srnk[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16, ylim=c(0,1),
         main = paste("LIU"," n:",n[i]),xlab="muestras",ylab="rank")
    abline(h=alpha,col=2)
    out<-(srnk[i,]<=alpha)+1
    points(x=1:n[i],srnk[i,],col=out)
    text(n[i]/2,0.5,labels = paste("LIU: LCI=",alpha,
                                   "- PR=",alpha),col = "black")

    plot(sdep[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,1),
         main = paste("MC"," n:",n[i]),xlab="muestras",ylab="profundidad")
    abline(h=round(lc.prmc$LCI,4), col="blue")
    out<-(sdep[i,]<=lc.prmc$LCI)+1
    points(x=1:n[i],sdep[i,],col=out)
    text(n[i]/2,0.5,labels = paste("MC: LCI=",round(lc.prmc$LCI,3),
                                   "- PR=",round(lc.prmc$RP,4)),col = "black")

    plot(sdep[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,1),
         main = paste("BOOT"," n:",n[i]),xlab="muestras",ylab="profundidad")
    abline(h=round(lc.prboot$LCI,4), col="brown")
    out<-(sdep[i,]<=lc.prboot$LCI)+1
    points(x=1:n[i],sdep[i,],col=out)
    text(n[i]/2,0.5,labels = paste("BOOT: LCI=",round(lc.prboot$LCI,3),
                                   "- PR=",round(lc.prboot$RP,4)),col = "black")

  }
}

#####
#Gráfico Variación aislada

grafvar<-function(res.sim,res.sima,n,t,mc=1000,alpha=0.05){

```

```

N<-length(n)
for (i in 1:N){
  #Variación aislada muestra completa
  sdep<-res.sima[[i]][[t[1]]]
  lcboot<-res.sim[[i]][[t[2]]];prvdep<-Lim(sdep,LCI = lcboot)$RP
  plot(sdep[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,1),
        main = paste("Muestra con Variación aislada ", "n:",n[i]),
        xlab="muestras",ylab="profundidad")
  abline(h=round(lcboot,4), col="blue")
  out<-(sdep[i,]<=lcboot)+1
  points(x=1:n[i],sdep[i,],col=out)
  text(n[i]/2,0.05,labels = paste("BOOT: LCI=",round(lcboot,3),
                                   "- PR=",round(prvdep,4)),col = "black")

  #Variación aislada
  va<-sdep[,n[i]];prva<-Lim(va,alpha=alpha,LCI = lcboot)$RP
  plot(va[1:100], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,1),
        main = paste("Variaciones aisladas para ", "n:",n[i]),
        xlab="muestras",ylab="profundidad")
  abline(h=round(lcboot,4), col="blue")
  out<-(va<=lcboot)+1
  points(x=1:mc,va,col=out)
  text(50,0.05,labels = paste("BOOT: LCI=",round(lcboot,3),
                                   "- PR=",round(prva,4)),col = "black")

  #Puntos originales sin variación
  cva<-sdep[,1:(n[i]-1)];prvca<-Lim(cva,alpha=alpha,LCI = lcboot)$RP
  plot(cva[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,1),
        main = paste("Muestra sin Variación aislada ", "n:",n[i]),
        xlab="muestras",ylab="profundidad")
  abline(h=round(lcboot,4), col="blue")
  out<-(cva[i,]<=lcboot)+1
  points(x=1:(n[i]-1),cva[i,],col=out)
  text(n[i]/2,0.05,labels = paste("BOOT: LCI=",round(lcboot,3),
                                   "- PR=",round(prvca,4)),col = "black")
}
}
#####

```

```

#Gráfico de control para Fase II
grafase2<-function(ha.sim,res.sim,n,delta,alpha=0.05,ndat=100){
N<-length(n);D<-length(delta)
  for (j in 1:D){
  for (i in 1:N){
    #Límite de control y porcentaje de rechazo
    LCI<-res.sim[[i]][[5]];PR <-ha.sim[[i]][[1]][j,9]
    mu<-ha.sim[[i]][[1]][j,1]
    sdep<-ha.sim[[i]][[3]][[j]][1,][1:ndat]
    plot(sdep, lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,1),
        main = paste("GC depth - ", "n=",n[i],
                    "mu=" ,mu),xlab="muestras",ylab="profundidad")
    abline(h=round(LCI,4), col="blue")
    out<-(sdep<=LCI)+1
    points(x=1:ndat,sdep,col=out)
    text(50,1,labels = paste("MC: LCI=",round(LCI,3),
                            "- PR=",round(PR,4)),col = "black")
  }
}
}
#####
#Extracción de datos de una lista
extrdat<-function(ha.sim,n,delta,alpha=0.05,ndat=100){
  N<-length(n);D<-length(delta)
  deps<-list(N*D)
  d<-1
  for (j in 1:D){
    for (i in 1:N){
      sdep<-ha.sim[[i]][[3]][[j]][1,][1:ndat]
      deps[[d]]<-sdep
      d<-d+1
    }
  }
  return(deps)
}

```

A.2.2. Simulación Fase I caso independiente

Para generar estudios de simulación para distintos tamaños de muestra se varía el vector de muestras

```
rm(list = ls())
memory.limit(4*10024^3)
#Para distintos valores de n se aplica el método Monte Carlo
#y bootstrap bajo hipótesis nula
mc<-1000
nboot<-500
alpha<-0.05
n <- c(25,50,100,15)
N<-length(n)
p<-2
q<-5
#Parámetro bootstrap suavizado
h<-0
mu<-c(1,2,3,4)
#####
####Hipótesis inicial
#Para n=25
dataref<-array(NA,dim=c(n[1],p,mc))
dep<-matrix(NA,nrow=mc,ncol=n[1],byrow=TRUE)
rank<-matrix(NA,nrow=mc,ncol=n[1],byrow=TRUE)
#Monte Carlo
simmc<-sim.mc(mc=mc,n=n[1],p=p,q=q,depend = FALSE,func = mdepth.SD)
dataref<-array(sapply(simmc,function(x) x$data),dim=c(n[1],p,mc))
dep<-matrix(sapply(simmc,function(x) x$dep),nrow=mc,ncol=n[1],byrow=TRUE)
rank<-matrix(sapply(simmc,function(x) x$rank),nrow=mc,ncol=n[1],byrow=TRUE)
lcmc<-Lim(dep,alpha = alpha)$LCI
#Bootstrap
bootdep<-bootker(dataref,mc=mc,nboot=nboot,h=h,data.ref=NULL,func = mdepth.SD)
denboot<-density(bootdep)
lcboot<-Lim(bootdep,alpha = alpha)$LCI
prboot<-Lim(bootdep,alpha = alpha)$RP
res_n1<-list(dataref,dep,denboot,lcmc,lcboot,rank,prboot)
```

```

n1<-n[1]
save(res_n1,n1,file="res_n1.Rdata")
####Hipótesis alternativa
#Añadiendo una variación aislada (punto atípico)
vvas<-numeric(N)
for (i in 1:N){
  vdep<-vais(dataref,n=n[1],mu=mu[i],mc=mc,func=mdepth.SD)$dep
  va<-vdep[,n[1]]
  vvas[i]<-Lim(va,alpha=alpha,LCI = lcboot)$RP
}
#Puntos originales sin variación
cva<-vdep[,1:(n[1]-1)]
Lim(cva,alpha=alpha,LCI = lcboot)$RP
#Porcentaje rechazo toda la muestra
prvdep<-Lim(vdep,alpha=alpha,LCI = lcboot)$RP
#Eliminando 10% de puntos
persim<-trdat(dataref,n=n[1],mc=mc,per=0.9,func = mdepth.SD)
perdat<-persim$data
perdep<-persim$dep
perrank<-persim$rank
##Bootstrap
dep.boot<-bootker(perdat,mc=mc,nboot=nboot,h=h,data.ref=NULL,func = mdepth.SD)
den.boot<-density(dep.boot)
per.lc<-Lim(dep.boot,alpha = alpha)$LCI
prper<-Lim(dep.boot,alpha = alpha)$RP
#Variación aislada muestra recortada
vvasr<-numeric(N)
for (i in 1:N){
  vrdep<-vais(perdat,n=n[1],mu=mu[i],mc=mc,func=mdepth.SD)$dep
  va<-vrdep[,n[1]]
  vvasr[i]<-Lim(va,alpha=alpha,LCI = lcboot)$RP
}
res_n1a<-list(vdep,perdep,prvdep,per.lc,prper,perrank,den.boot,vrdep,vvas,vvasr)

```

A.2.3. Simulación Fase I caso dependiente

La simulación para la Fase I para observaciones autocorrelacionadas se realiza para distintos valores de muestra y coeficiente de correlación ρ .

```
rm(list = ls())
memory.limit(4*2024^3)
source("estadísticos_SD.R")
#Para distintos valores de n se aplica el método Monte Carlo
#y bootstrap bajo hipótesis nula
mc<-1000
nboot<-500
alpha<-0.05
n <- c(25,50,100,15)
N<-length(n)
p<-2
q<-5
ro<-0.7
mu<-c(1,2,3,4)
l<-5 #media distribución geométrica
#####
####Hipótesis inicial
#Para n=25
dataref<-array(NA,dim=c(n[1],p,mc))
dep<-matrix(NA,nrow=mc,ncol=n[1],byrow=TRUE)
rank<-matrix(NA,nrow=mc,ncol=n[1],byrow=TRUE)
#Monte Carlo
simmc<-sim.mc(mc=mc,n=n[1],p=p,q=q,ro=ro,depend = TRUE,func = mdepth.SD)
dataref<-array(sapply(simmc,function(x) x$data),dim=c(n[1],p,mc))
dep<-matrix(sapply(simmc,function(x) x$dep),nrow=mc,ncol=n[1],byrow=TRUE)
rank<-matrix(sapply(simmc,function(x) x$rank),nrow=mc,ncol=n[1],byrow=TRUE)
lcmc<-Lim(dep,alpha = alpha)$LCI
#Bootstrap
bootdep<-bootmc(ni=n[1],data=dataref,mc=mc,nboot=nboot,data.ref=NULL,depend=TRUE,
               stat="dep",statistic=tsrank,l=1)
denboot<-density(bootdep)
lcboot<-Lim(bootdep,alpha = alpha)$LCI
```

```

prboot<-Lim(bootdep,alpha = alpha)$RP
res_n1<-list(dataref,dep,denboot,lcmc,lcboot,rank,prboot)
n1<-n[1]
save(res_n1,n1,file="res_n1.Rdata")
####Hipótesis alternativa
#Añadiendo una variación aislada (punto atípico)
vvas<-numeric(N)
for (i in 1:N){
  vdep<-vais(dataref,n=n[1],mu=mu[i],mc=mc,func=mdepth.SD,ro=ro)$dep
  va<-vdep[,n[1]]
  vvas[i]<-Lim(va,alpha=alpha,LCI = lcboot)$RP
}
#Puntos originales sin variación
cva<-vdep[,1:(n[1]-1)]
Lim(cva,alpha=alpha,LCI = lcboot)$RP
#Porcentaje rechazo toda la muestra
prvdep<-Lim(vdep,alpha=alpha,LCI = lcboot)$RP
#Eliminando 10% de puntos
persim<-trdat(dataref,n=n[1],mc=mc,per=0.9,func = mdepth.SD)
perdat<-persim$data
perdep<-persim$dep
perrank<-persim$rank
##Bootstrap
dep.boot<-bootmc(ni=n[1],data=perdat,mc=mc,nboot=nboot,data.ref=NULL,depend=TRUE,
                 stat="dep",statistic=tsrank,l=1)
den.boot<-density(dep.boot)
per.lc<-Lim(dep.boot,alpha = alpha)$LCI
prper<-Lim(dep.boot,alpha = alpha)$RP
#Variación aislada muestra recortada
vvasr<-numeric(N)
for (i in 1:N){
  vrdep<-vais(perdat,n=n[1],mu=mu[i],mc=mc,func=mdepth.MhD,ro=ro)$dep
  va<-vrdep[,n[1]]
  vvasr[i]<-Lim(va,alpha=alpha,LCI = lcboot)$RP
}

```

```
res_n1a<-list(vdep,perdep,prvdep,per.lc,prper,perrank,den.boot,vrdep,vvas,vvasr)
```

Generación de tablas para Fase I

```
rm(list=ls())
library(knitr)
res.sim <- list(res_n1,res_n2,res_n3,res_n4)
res.sima <- list(res_n1a,res_n2a,res_n3a,res_n4a)
n <- c(n1,n2,n3,n4)
N <- length(n)
alpha<-0.05
table<-data.frame(n = rep(n),LCImc=numeric(N),LCIboot=numeric(N),
                  LCIRank=numeric(N),PRmc=numeric(N),PRboot=numeric(N),
                  PRRank=numeric(N),PRvar=numeric(N))
tablerec<-data.frame(n = rep(n),LCImc=numeric(N),LCIboot=numeric(N),
                    LCIRank=numeric(N),PRmc=numeric(N),PRboot=numeric(N),
                    PRRank=numeric(N),PRvar=numeric(N))
tableha<-data.frame(n = rep(n),mu1=numeric(N),mu2=numeric(N),
                   mu3=numeric(N),mu4=numeric(N),mu1rm=numeric(N),mu2rm=numeric(N),
                   mu3rm=numeric(N),mu4rm=numeric(N))
ncol(tableha)
for (i in 1:N) {
  dep<-res.sim[[i]][[2]];lc.prmc<-Lim(dep,alpha = alpha)
  lc.boot<-res.sim[[i]][[5]];prboot<-res.sim[[i]][[7]]
  vdep<-res.sima[[i]][[1]];lc.prv<-Lim(vdep,alpha = alpha,LCI=lc.boot)
  table[i,2]<-lc.prmc$LCI;table[i,3]<-lc.boot
  table[i,4]<-alpha;table[i,5]<-lc.prmc$RP
  table[i,6]<-prboot;table[i,7]<-alpha;table[i,8]<-lc.prv$RP
}

for (i in 1:N) {
  dep<-res.sima[[i]][[2]];lc.prmc<-Lim(dep,alpha = alpha)
  lc.boot<-res.sima[[i]][[4]];prboot<-res.sima[[i]][[5]]
  vdep<-res.sima[[i]][[8]];lc.prv<-Lim(vdep,alpha = alpha,LCI=lc.boot)
  tablerec[i,2]<-lc.prmc$LCI;tablerec[i,3]<-lc.boot
  tablerec[i,4]<-alpha;tablerec[i,5]<-lc.prmc$RP
  tablerec[i,6]<-prboot;tablerec[i,7]<-alpha;tablerec[i,8]<-lc.prv$RP
}
```

```

}
for (i in 1:N) {
  tableha[1,i+1]<-res.sima[[1]][[9]][[i]]
  tableha[2,i+1]<-res.sima[[2]][[9]][[i]]
  tableha[3,i+1]<-res.sima[[3]][[9]][[i]]
  tableha[4,i+1]<-res.sima[[4]][[9]][[i]]
  tableha[1,i+5]<-res.sima[[1]][[10]][[i]]
  tableha[2,i+5]<-res.sima[[2]][[10]][[i]]
  tableha[3,i+5]<-res.sima[[3]][[10]][[i]]
  tableha[4,i+5]<-res.sima[[4]][[10]][[i]]
}
tb<-list(table,tablerec,tableha)

```

A.2.4. Simulación Fase II caso independiente

Caso para $n = 25$ el código para distintos tamaños de muestra es similar.

```

rm(list=ls())

library(knitr)
#Límite de control MC
LCI<-res_n1[[4]]
#Límite de control boot
LCIboot<-res_n1[[5]]
#' ## Análisis de escenarios
#' ### Variación de la media y ro
n <- c(n1,n2,n3,n4)
N <- length(n)
m <- 1000 #muestra referencia
M <- length(m)
p <- 2
q <- 5
delta<-c(-3.2,-2.4,-1.6,-0.8,0,0.8,1.6,2.4,3.2)
U <- length(delta)
mu_0 <- rep(0,p)
mu <- matrix(NA,nrow = p,ncol = U)
for (id in 1:U) {

```

```

    mu[,id] <- c(mu_0[1]+delta[id],mu_0[-1])
  }
mu2 <- matrix(NA,nrow = p,ncol = U)
for (id in 1:U) {
  mu2[,id] <- c(mu_0[1]+delta[id],mu_0[1]+delta[id])
}
# Generación de muestras para profundidad y rank bajo hipótesis alternativa
#para distintos valores de m
#####
#Variación de una componente del vector de medias
mc<-1000
alfa<-0.05
t25<-sim.esc(ni=25,m=m,mu=mu,mc=mc,dateref = res_n1[[1]],statistic = "dep",
             depend = FALSE,func = mdepth.SD)
ldep25<-t25$dep
lrank25<-t25$rank
#Porcentaje de rechazo
#MC
prmc<-PPnew(ldep25,LCIt = LCI)
#boot
prboot<-PPnew(ldep25,LCIt = LCIboot)
#rank
LCIt<-alfa
prliu<-PPnew(lrank25,LCIt =LCIt)

table25<-data.frame(mu = rep(delta, each=M),m = rep(rep(m,M),M),
                    n = rep(rep(n1,M),U),LCIliu= numeric(M*U),LCImc = numeric(M*U),
                    LCIboot = numeric(M*U),PRliu=numeric(M*U),PRmc=numeric(M*U),
                    PRboot=numeric(M*U))
table25[,4]<-LCIt
table25[,5]<-LCI
table25[,6]<-LCIboot
table25[,7]<-prliu
table25[,8]<-prmc
table25[,9]<-prboot
tb25<-table25

```

```

tb25[tb25[,1]==0,7:9]<-res_n1[[7]]

# Comparación porcentaje de rechazo Monte Carlo y bootstrap
#n=25
grafn1<-comp(tb25,ni=25,depend = FALSE,delta=delta)
# ARL_1
arl25<-data.frame(mu = rep(delta, each=M),ARLliu=numeric(M*U),ARLmc=numeric(M*U),
                  ARLboot=numeric(M*U))
arl25[,2]<-1/tb25[,7]
arl25[,3]<-1/tb25[,8]
arl25[,4]<-1/tb25[,9]

ha_n1<-list(tb25,grafn1,ldep25,arl25)

#####
#Variación de todos los componentes del vector de medias
mc<-1000
alfa<-0.05
t25<-sim.esc(ni=25,m=m,mu=mu2,mc=mc,dateref = res_n1[[1]],statistic = "dep",
            depend = FALSE,func = mdepth.SD)

ldep25<-t25$dep
lrank25<-t25$rank
#Porcentaje de rechazo
#MC
prmc<-PPnew(ldep25,LCIt = LCI)
#boot
prboot<-PPnew(ldep25,LCIt = LCIboot)
#rank
LCIt<-alfa
prliu<-PPnew(lrank25,LCIt =LCIt)

table25<-data.frame(mu = rep(delta, each=M),m = rep(rep(m,M),M),
                   n = rep(rep(n1,M),U),LCIliu= numeric(M*U),LCImc = numeric(M*U),
                   LCIboot = numeric(M*U),PRliu=numeric(M*U),PRmc=numeric(M*U),

```

```

PRboot=numeric(M*U)

table25[,4]<-LCIt
table25[,5]<-LCI
table25[,6]<-LCIboot
table25[,7]<-prliu
table25[,8]<-prmc
table25[,9]<-prboot
tb25<-table25
tb25[tb25[,1]==0,7:9]<-res_n1[[7]]

# Comparación porcentaje de rechazo Monte Carlo y bootstrap
#n=25
grafn1<-comp(tb25,ni=25,depend = FALSE,delta=delta)
# ARL_1
arl25<-data.frame(mu = rep(delta, each=M),ARLliu=numeric(M*U),ARLmc=numeric(M*U),
                  ARLboot=numeric(M*U))
arl25[,2]<-1/tb25[,7]
arl25[,3]<-1/tb25[,8]
arl25[,4]<-1/tb25[,9]

ha_n1a<-list(tb25,grafn1,ldep25,arl25)

```

A.2.5. Simulación Fase II caso dependiente

Caso para $n = 25$, el código para distintos tamaños de muestra es similar, la diferencia con el caso independiente radica en el coeficiente de correlación.

```

rm(list=ls())
library(knitr)
#Límite de control MC
LCI<-res_n1[[4]]
#Límite de control boot
LCIboot<-res_n1[[5]]

#' ## Análisis de escenarios
#' ### Variación de la media y ro
n <- c(n1,n2,n3)

```

```

N <- length(n)
m <- 1000
M <- length(m)
ro <- c(0.5,0.7,0.9)
K <- length(ro)
p <- 2
q <- 5
delta<-c(-3.2,-2.4,-1.6,-0.8,0,0.8,1.6,2.4,3.2)
U <- length(delta)
mu_0 <- rep(0,p)
mu <- matrix(NA,nrow = p,ncol = U)
for (id in 1:U) {
  mu[,id] <- c(mu_0[1]+delta[id],mu_0[-1])
}
mu2 <- matrix(NA,nrow = p,ncol = U)
for (id in 1:U) {
  mu2[,id] <- c(mu_0[1]+delta[id],mu_0[1]+delta[id])
}
# Generación de muestras para profundidad y rank bajo hipótesis alternativa
#para distintos valores de m
#####
#Variación de una componente del vector de medias
mc<-1000
alfa<-0.05
t25<-sim.esc(ni=25,m=m,mu=mu,ro=ro,mc=mc,dateref = res_n1[[1]],statistic = "dep",
depend = TRUE,func = mdepth.SD)
ldep25<-t25$dep
lrank25<-t25$rank
#Porcentaje de rechazo
#MC
prmc<-PP(ldep25,LCI = LCI,depend = TRUE)
#boot
prboot<-PP(ldep25,LCI = LCIboot,depend=TRUE)
#rank
LCIr<-alfa
prliu<-PP(lrank25,LCI =LCIr,depend=TRUE)

```

```

table25<-data.frame(mu = rep(delta, each=M*K),ro = rep(rep(ro, each=M),U),
                    m = rep(rep(m,M),U),
                    n = rep(rep(n1,M*K),U),
                    LCliiu= numeric(M*K*U),
                    LCImc = numeric(M*K*U),
                    LCiboot = numeric(M*K*U),PRliu=numeric(M*K*U),
                    PRmc=numeric(M*K*U),
                    PRboot=numeric(M*K*U))

table25[,5]<-LCIr
table25[,6]<-LCI
table25[,7]<-LCiboot
table25[,8]<-prliu
table25[,9]<-prmc
table25[,10]<-prboot
tb25<-table25
tb25[tb25[,1]==0,8:10]<-res_n1[[7]]

# Comparación porcentaje de rechazo Monte Carlo y bootstrap
#n=25
grafn1<-comp(tb25,ni=25,depend = TRUE,delta=delta)
# ARL_1
arl25<-data.frame(mu = rep(delta, each=M*K),ro = rep(rep(ro, each=M),U),
                  ARLliu=numeric(M*U*K),ARLmc=numeric(M*U*K),
                  ARLboot=numeric(M*U*K))

arl25[,3]<-1/tb25[,8]
arl25[,4]<-1/tb25[,9]
arl25[,5]<-1/tb25[,10]

ha_n1<-list(tb25,grafn1,ldep25,arl25)

#####
#Variación de todos los componentes del vector de medias
mc<-1000
alfa<-0.05

```

```

t25<-sim.esc(ni=25,m=m,mu=mu2,ro=ro,mc=mc,dateref = res_n1[[1]],statistic = "dep",
depend = TRUE,func = mdepth.SD)
ldep25<-t25$dep
lrank25<-t25$rank
#Porcentaje de rechazo
#MC
prmc<-PP(ldep25,LCI = LCI,depend = TRUE)
#boot
prboot<-PP(ldep25,LCI = LCIboot,depend=TRUE)
#rank
LCIr<-alfa
prliu<-PP(lrank25,LCI =LCIr,depend=TRUE)

table25<-data.frame(mu = rep(delta, each=M*K),ro = rep(rep(ro, each=M),U),
                    m = rep(rep(m,M),U),
                    n = rep(rep(n1,M*K),U),LCIliu= numeric(M*K*U),
                    LCImc = numeric(M*K*U),
                    LCIboot = numeric(M*K*U),PRliu=numeric(M*K*U),
                    PRmc=numeric(M*K*U),
                    PRboot=numeric(M*K*U))

table25[,5]<-LCIr
table25[,6]<-LCI
table25[,7]<-LCIboot
table25[,8]<-prliu
table25[,9]<-prmc
table25[,10]<-prboot
tb25<-table25
tb25[tb25[,1]==0,8:10]<-res_n1[[7]]

# Comparación porcentaje de rechazo Monte Carlo y bootstrap
#n=25
grafn1<-comp(tb25,ni=25,depend = TRUE,delta=delta)
# ARL_1
arl25<-data.frame(mu = rep(delta, each=M*K),ro = rep(rep(ro, each=M),U),
                  ARLliu=numeric(M*U*K),ARLmc=numeric(M*U*K),
                  ARLboot=numeric(M*U*K))

```

```

arl25[,3]<-1/tb25[,8]
arl25[,4]<-1/tb25[,9]
arl25[,5]<-1/tb25[,10]

```

```

ha_n1a<-list(tb25,grafn1,ldep25,arl25)

```

A.2.6. Tablas para Fase II, variación de un componente del vector de medias

```

rm(list=ls())

```

```

res.sim <- list(res_n1,res_n2,res_n3,res_n4)
ha.sim <- list(ha_n1,ha_n2,ha_n3,ha_n4)
delta<-c(-3.2,-2.4,-1.6,-0.8,0,0.8,1.6,2.4,3.2)
n<-c(n1,n2,n3)
par(mfrow=c(1,3))
##Gráfico control de calidad
grafase2(ha.sim=ha.sim,res.sim=res.sim,n=n,delta=delta)
par(mfrow=c(1,1))

```

```

#Extracción de datos

```

```

extrha1<-extrdat(ha.sim=ha.sim,n=n,delta=delta)
##Comparación porcentaje de rechazo
n<-c(n1,n2,n3,n4)
comp(ha.sim[[4]][[1]],ni=15,depend = FALSE,delta=delta)
comp(ha.sim[[1]][[1]],ni=25,depend = FALSE,delta=delta)
comp(ha.sim[[2]][[1]],ni=50,depend = FALSE,delta=delta)
comp(ha.sim[[3]][[1]],ni=100,depend = FALSE,delta=delta)

```

```

#Tabla porcentaje de rechazo y límites de control

```

```

###n=15
ha.sim[[4]][[1]]
###n=25
ha.sim[[1]][[1]]
###n=50
ha.sim[[2]][[1]]

```

```

###n=100
ha.sim[[3]][[1]]

tbpr1<-list(ha.sim[[1]][[1]],ha.sim[[2]][[1]],ha.sim[[3]][[1]],
            ha.sim[[4]][[1]])

#Average Run Length
#n=15
ha.sim[[4]][[4]]
#n=25
ha.sim[[1]][[4]]
#n=50
ha.sim[[2]][[4]]
#n=100
ha.sim[[3]][[4]]

arl1<-list(ha.sim[[4]][[4]],ha.sim[[1]][[4]],ha.sim[[2]][[4]],
           ha.sim[[3]][[4]])

```

A.2.7. Tablas para Fase II, variación todos los componentes del vector de medias

```

rm(list=ls())

res.sima <- list(res_n1a,res_n2a,res_n3a,res_n4a)
ha.sima <- list(ha_n1a,ha_n2a,ha_n3a,ha_n4a)
delta<-c(-3.2,-2.4,-1.6,-0.8,0,0.8,1.6,2.4,3.2)
n<-c(n1,n2,n3)
par(mfrow=c(1,3))
##Gráfico control de calidad
grafase2(ha.sim=ha.sima,res.sim=res.sima,n=n,delta=delta)
par(mfrow=c(1,1))

#Extracción de datos
extrha2<-extrdat(ha.sim=ha.sima,n=n,delta=delta)

```

```

##Comparación porcentaje de rechazo
n<-c(n1,n2,n3,n4)
comp(ha.sima[[4]][[1]],ni=15,depend = FALSE,delta=delta)
comp(ha.sima[[1]][[1]],ni=25,depend = FALSE,delta=delta)
comp(ha.sima[[2]][[1]],ni=50,depend = FALSE,delta=delta)
comp(ha.sima[[3]][[1]],ni=100,depend = FALSE,delta=delta)

#Tabla porcentaje de rechazo y límites de control
###n=15
ha.sima[[4]][[1]]
###n=25
ha.sima[[1]][[1]]
###n=50
ha.sima[[2]][[1]]
###n=100
ha.sima[[3]][[1]]

tbpr2<-list(ha.sima[[1]][[1]],ha.sima[[2]][[1]],ha.sima[[3]][[1]],
            ha.sima[[4]][[1]])

#Average Run Length
#n=15
ha.sima[[4]][[4]]
#n=25
ha.sima[[1]][[4]]
#n=50
ha.sima[[2]][[4]]
#n=100
ha.sima[[3]][[4]]

arl2<-list(ha.sima[[4]][[4]],ha.sima[[1]][[4]],ha.sima[[2]][[4]]
           ,ha.sima[[3]][[4]])

```

A.2.8. Gráficos de control para Fases I y II

Fase I

```
#Hipótesis nula
n<-c(n1,n2,n3,n4)
N<-length(n)
par(mfrow=c(1,3))
for (i in 1:N){
  sdep<-res.sim[[i]][[2]]; lc.prmc<-Lim(sdep,alpha = alpha)
  srnk<-res.sim[[i]][[6]]; lc.boot<-res.sim[[i]][[4]]
  bootdep<-res.sim[[i]][[3]]; pr.boot<-res.sim[[i]][[5]]
  ### Densidad
  plot(density(sdep),main = paste("n= ",n[i],"alpha= ",alpha),xlab="profundidad",
       ylim=c(0,10),col=2)
  lines(bootdep,main = paste("n= ",n[i],"alpha= ",alpha),xlab="profundidad",
       ylim=c(0,10),col=3)
  legend("topleft",legend = c(paste("MC"),paste("Bootstrap")), bty = "n",
       col = c(2,3),lty = 1)
}
par(mfrow=c(1,3))
for (i in 1:N){
  ##### Grafico de control de calidad
  sdep<-res.sim[[i]][[2]]; lc.prmc<-Lim(sdep,alpha = alpha)
  srnk<-res.sim[[i]][[6]]
  lc.boot<-res.sim[[i]][[5]]; pr.boot<-res.sim[[i]][[7]]
  plot(srnk[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16, ylim=c(0,1),
       main = paste("LIU", " n:",n[i]),xlab="muestras",ylab="rank")
  abline(h=alpha,col=2)
  out<-(srnk[i,]<=alpha)+1
  points(x=1:n[i],srnk[i,],col=out)
  text(n[i]/2,0.5,labels = paste("LIU: LCI=",alpha,
                                "- PR=",alpha),col = "black")

  plot(sdep[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,0.4),
       main = paste("MC", " n:",n[i]),xlab="muestras",ylab="profundidad")
  abline(h=round(lc.prmc$LCI,4), col="blue")
}
```

```

out<-(sdep[i,]<=lc.prmc$LCI)+1
points(x=1:n[i],sdep[i,],col=out)
text(n[i]/2,0.4,labels = paste("MC: LCI=",round(lc.prmc$LCI,3),
                                "- PR=",round(lc.prmc$RP,4)),col = "black")

plot(sdep[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,0.4),
      main = paste("BOOT", " n:",n[i]),xlab="muestras",ylab="profundidad")
abline(h=round(lc.boot,4), col="brown")
out<-(sdep[i,]<=lc.boot)+1
points(x=1:n[i],sdep[i,],col=out)
text(n[i]/2,0.4,labels = paste("BOOT: LCI=",round(lc.boot,3),
                                "- PR=",round(pr.boot,4)),col = "black")

}
par(mfrow=c(1,1))
#Hipótesis alternativa
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
for (i in 1:N){
  #Variación aislada muestra completa
  sdep<-res.sima[[i]][[1]]
  prvdep<-res.sima[[i]][[3]];lcboot<-res.sim[[i]][[5]]
  plot(sdep[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,0.5),
        main = paste("Muestra con Variación aislada ", "n:",n[i]),xlab="muestras",
                    ylab="profundidad")
  abline(h=round(lcboot,4), col="blue")
  out<-(sdep[i,]<=lcboot)+1
  points(x=1:n[i],sdep[i,],col=out)
  text(n[i]/2,0.45,labels = paste("BOOT: LCI=",round(lcboot,3),
                                  "- PR=",round(prvdep,4)),col = "black")

  #Variación aislada
  va<-sdep[,n[i]];prva<-Lim(va,alpha=alpha,LCI = lcboot)$RP
  plot(va[1:100], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,0.5),
        main = paste("Variaciones aislada para ", "n:",n[i]),xlab="muestras",
                    ylab="profundidad")
  abline(h=round(lcboot,4), col="blue")
  out<-(va<=lcboot)+1

```

```

points(x=1:mc,va,col=out)
text(50,0.45,labels = paste("BOOT: LCI=",round(lcboot,3),
                             "- PR=",round(prva,4)),col = "black")

#Puntos originales sin variación
cva<-sdep[,1:(n[i]-1)];prvca<-Lim(cva,alpha=alpha,LCI = lcboot)$RP
plot(cva[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,0.55),
     main = paste("Muestra sin Variación aislada ", "n:",n[i]),xlab="muestras",
     ylab="profundidad")
abline(h=round(lcboot,4), col="blue")
out<-(cva[i,]<=lcboot)+1
points(x=1:(n[i]-1),cva[i,],col=out)
text(n[i]/2,0.5,labels = paste("BOOT: LCI=",round(lcboot,3),
                              "- PR=",round(prvca,4)),col = "black")
}
par(mfrow=c(1,1))
#Recorte de muestra
par(mfrow=c(1,3))
for (i in 1:N){
sdep<-res.sima[[i]][[2]]; lc.prmc<-Lim(sdep,alpha = alpha)
srnk<-res.sima[[i]][[6]]; lc.boot<-res.sima[[i]][[5]]
bootdep<-res.sima[[i]][[7]]; pr.boot<-res.sima[[i]][[7]]
### Densidad
plot(density(sdep),main = paste("n= ",n[i],"alpha= ",alpha),xlab="profundidad",
     ylim=c(0,12),col=2)
lines(bootdep,main = paste("n= ",n[i],"alpha= ",alpha),xlab="profundidad",
     ylim=c(0,12),col=3)
legend("topleft",legend = c(paste("MC"),paste("Bootstrap")), bty = "n",
     col = c(2,3),lty = 1)
par(mfrow=c(1,3))
for (i in 1:N){
##### Grafico de control de calidad
sdep<-res.sima[[i]][[2]]; lc.prmc<-Lim(sdep,alpha = alpha)
srnk<-res.sima[[i]][[6]]; lc.boot<-res.sima[[i]][[4]]
pr.boot<-res.sima[[i]][[5]]
plot(srnk[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16, ylim=c(0,1),
     main = paste("LIU", " n:",n[i]),xlab="muestras",ylab="rank")

```

```

abline(h=alpha,col=2)
out<-(srank[i,]<=alpha)+1
points(x=1:n[i],srank[i,],col=out)
text(n[i]/2,0.5,labels = paste("LIU: LCI=",alpha,
                                "- PR=",alpha),col = "black")

plot(sdep[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,0.4),
      main = paste("MC"," n:",n[i]),xlab="muestras",ylab="profundidad")
abline(h=round(lc.prmc$LCI,4), col="blue")
out<-(sdep[i,]<=lc.prmc$LCI)+1
points(x=1:n[i],sdep[i,],col=out)
text(n[i]/2,0.4,labels = paste("MC: LCI=",round(lc.prmc$LCI,3),
                                "- PR=",round(lc.prmc$RP,4)),col = "black")

plot(sdep[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,0.4),
      main = paste("BOOT"," n:",n[i]),xlab="muestras",ylab="profundidad")
abline(h=round(lc.boot,4), col="brown")
out<-(sdep[i,]<=lc.boot)+1
points(x=1:n[i],sdep[i,],col=out)
text(n[i]/2,0.4,labels = paste("BOOT: LCI=",round(lc.boot,3),
                                "- PR=",round(pr.boot,4)),col = "black")

}
par(mfrow=c(1,1))
#Hipótesis alternativa recorte de muestra
layout(matrix(c(1,1,2,3), 2, 2, byrow = TRUE))
for (i in 1:N){
  #Variación aislada muestra completa
  sdep<-res.sima[[i]][[8]]
  lcboot<-res.sim[[i]][[5]];prvdep<-Lim(sdep,LCI = lcboot)$RP
  plot(sdep[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,0.5),
        main = paste("Muestra con Variación aislada ", "n:",n[i]),xlab="muestras",
                    ylab="profundidad")
  abline(h=round(lcboot,4), col="blue")
  out<-(sdep[i,]<=lcboot)+1
  points(x=1:n[i],sdep[i,],col=out)
}

```

```

text(n[i]/2,0.45,labels = paste("BOOT: LCI=",round(lcboot,3),
                                "- PR=",round(prvdep,4)),col = "black")
#Variación aislada
va<-sdep[,n[i]];prva<-Lim(va,alpha=alpha,LCI = lcboot)$RP
plot(va[1:100], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,0.5),
      main = paste("Variaciones aislada para ", "n:",n[i]),xlab="muestras",
      ylab="profundidad")
abline(h=round(lcboot,4), col="blue")
out<-(va<=lcboot)+1
points(x=1:mc,va,col=out)
text(50,0.45,labels = paste("BOOT: LCI=",round(lcboot,3),
                              "- PR=",round(prva,4)),col = "black")
#Puntos originales sin variación
cva<-sdep[,1:(n[i]-1)];prvca<-Lim(cva,alpha=alpha,LCI = lcboot)$RP
plot(cva[i,], lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,0.5),
      main = paste("Muestra sin Variación aislada ", "n:",n[i]),xlab="muestras",
      ylab="profundidad")
abline(h=round(lcboot,4), col="blue")
out<-(cva[i,]<=lcboot)+1
points(x=1:(n[i]-1),cva[i,],col=out)
text(n[i]/2,0.45,labels = paste("BOOT: LCI=",round(lcboot,3),
                              "- PR=",round(prvca,4)),col = "black")

```

Fase II

```

#Variación una componente vector de medias
m<-1000
M<-length(m)
D <- length(delta)
ndat<-100
par(mfrow=c(1,3))
n<-c(n1,n2,n3)
N<-length(n)
d<-1
for (j in 1:D){
  for (i in 1:N){
    #Límite de control y porcentaje de rechazo

```

```

LCI<-res.sim[[i]][[5]];PR <-tbpr1[[i]][j,9]
mu<-tbpr1[[i]][j,1]
sdep<-extrha1[[d]]
plot(sdep, lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,0.4),
      main = paste("GC depth - ", "n=",n[i],
                   "mu=" ,mu),xlab="muestras",ylab="profundidad")
abline(h=round(LCI,4), col="blue")
out<-(sdep<=LCI)+1
points(x=1:ndat,sdep,col=out)
text(50,0.35,labels = paste("MC: LCI=",round(LCI,3),
                             "- PR=",round(PR,4)),col = "black")

d<-d+1
}
}
par(mfrow=c(1,1))
#Variación de dos componentes vector de medias
D <- length(delta)
ndat<-100
par(mfrow=c(1,3))
n<-c(n1,n2,n3)
N<-length(n)
d<-1
for (j in 1:D){
  for (i in 1:N){
    #Límite de control y porcentaje de rechazo
    LCI<-res.sima[[i]][[5]];PR <-tbpr2[[i]][j,9]
    mu<-tbpr2[[i]][j,1]
    sdep<-extrha2[[d]]
    plot(sdep, lwd = 0.5, lty = 2, type="b",col = 2,pch=16,ylim=c(0,1),
          main = paste("GC depth - ", "n=",n[i],
                       "mu=" ,mu),xlab="muestras",ylab="profundidad")
    abline(h=round(LCI,4), col="blue")
    out<-(sdep<=LCI)+1
    points(x=1:ndat,sdep,col=out)
    text(50,1,labels = paste("MC: LCI=",round(LCI,3),
                              "- PR=",round(PR,4)),col = "black")
  }
}

```

```

    d<-d+1
  }
}
par(mfrow=c(1,1))

```

A.2.9. Aplicación a caso de estudio extrusión de palitos, caso independiente

```

library(qcr)
source("estadísticos_SD.R")
library(readxl)
library(xtable)
library(TSA)
base <- read_excel("Data.xlsx",sheet="Data",na="",col_names = TRUE)
str(base)
ro<-cor(base$Diámetro[1:90],base$Longitud[1:90])
nboot<-500
alpha<-0.05
bas<-cbind(base$Diámetro[1:90],base$Longitud[1:90])
mvShapiro.Test(bas)
dat<-cbind(base$Diámetro[1:60],base$Longitud[1:60])
datref<-cbind(base$Diámetro[61:90],base$Longitud[61:90])
n<-nrow(dat)
m0<-nrow(datref)

bootapl<-function(data,nboot,h=0.001,data.ref=NULL,
                  func=func){
  n<-nrow(data)
  rs<-matrix(NA,ncol=n,nrow=nboot)
  rss<-numeric(n)
  for (k in 1:nboot)
  {
    if(!is.null(data.ref)){
      dep.ref<-func(data)$dep
      dep<-func(data.ref,data)$dep
      rank <- sapply(dep,function(x) mean(dep.ref <= x))

```

```

    }
    else{
      dep.ref<-func(data)$dep
      dep<-func(data)$dep
      rank <- sapply(dep,function(x) mean(dep.ref <= x))
    }
    rss<-sample(dep,n,replace=T)
    rs[k,]<-rss+h*rnorm(n,0,1)
  }
  return(rs)
}

###Gráfico de control estándar  $T^2$  Hotelling
#FASE I
data.mqcd <- mqcd(dat)
res.mqcs <- mqcs.t2(data.mqcd)
summary(res.mqcs)
plot(res.mqcs, title ="MCUSUM control chart for dowel pin")
lc<-res.mqcs$limits[2]
#FASE II
dowl<-mqcd(datref)
res.mqcs2 <- mqcs.t2(dowl)
mean(res.mqcs2$statistics>lc)
#Bootstrap muestra de monitoreo (FASE I)
datn<-dat[-c(35,43,49,53,54),]
n<-nrow(datn)
#n<-nrow(dat)
btdep<-bootapl(data=datn,nboot = nboot,func=mdepth.SD)
lcboot<-Lim(btdep,alpha = alpha)$LCI
#Gráfico de control profundidad (FASE II)
depth.x1<-mdepth.SD(datn)$dep
depth.x2<-mdepth.SD(datref,dat)$dep
depth.x<-c(depth.x1,depth.x2)
g = c(rep(2, n), rep(3, m0))
plot(depth.x1, lwd = 0.5, lty = 2, type="b",pch=16,col="green",
      main="Fase I observaciones independientes",xlab="muestras",ylab="profundidad")
abline(h=lcboot,col=4)

```

```

out<-(depth.x1<=lcboot)+1
points(x=1:n,depth.x1,col=out)
#Gráfico de control profundidad muestra monitoreo y referencia (FASE II)
plot(depth.x, lwd = 0.5, lty = 2, type="b",col = g,pch=16,
      main="Gráfico control profundidad",xlab="muestras",ylab="profundidad")
abline(h=lcboot,col=1)
out<-(depth.x2<=lcboot)+1
points(x=(n+1):(m0+n),depth.x2,col=out)
#Bootstrap muestra de referencia
btdepref<-bootapl(data=dat,nboot = nboot,data.ref = datref,func = mdepth.SD)
#porcentaje de rechazo profundidad
Lim(btdepref,LCI=lcboot)$RP

#####
#simulando observaciones de una distribución normal
shapiro.test(base$Diámetro)
shapiro.test(base$Longitud)
mu1<-mean(base$Diámetro)
mu2<-mean(base$Longitud)
mu<-c(mu1,mu2)
var1<-sd(base$Diámetro)^2
var2<-sd(base$Longitud)^2
cov1<-cov(base$Diámetro,base$Longitud)
cov2<-cov(base$Diámetro,base$Longitud)
Sigma<-matrix(c(var1,cov1,cov2,var2),2,2)

#Variación en la media para Fase II GC estándar
set.seed(123)
vargc<-mqcd(datref)
fas2dep<-mqcs.t2(vargc)
#Hacer cambios en la media desde 8 con saltos de 1
prGC<-numeric(5)
for (i in 1:5){
  fas2stat<-fas2dep$statistics+matrix(rep(2*i,times=30),nrow=30,byrow = TRUE)
  prGC[i]<-mean(fas2stat>lc)
}

```

```
prGC
```

```
#Variación en la media para Fase II medida de profundidad
#Variación una componente
PR<-numeric(5)
for (i in 1:5) {
  mualt<-c(mu1+i/15,mu2)
  dataref<-mvrnorm(n=30,mu=mualt,Sigma = Sigma)
  btdeprefvar<-bootapl(data=dat,nboot = nboot,data.ref = dataref,func=mdepth.SD)
  #porcentaje de rechazo
  PR[i]<-Lim(btdeprefvar,LCI=lcboot)$RP
}
PR
#Variación dos componentes
PR2<-numeric(5)
for (i in 1:5) {
  mualt<-c(mu1+i/15,mu2+i/100)
  dataref<-mvrnorm(n=30,mu=mualt,Sigma = Sigma)
  btdeprefvar<-bootapl(data=dat,nboot = nboot,data.ref = dataref,func=mdepth.SD)
  #porcentaje de rechazo
  PR2[i]<-Lim(btdeprefvar,LCI=lcboot)$RP
}
PR2
```

A.2.10. Aplicación caso de estudio extrusión de palitos, caso dependiente

```
library(qcr)
source("estadísticos_SD.R")
library(readxl)
library(xtable)
library(TSA)
base <- read_excel("Datos Dependientes.xlsx",sheet="Data",na="",col_names = TRUE)
str(base)
#correlación
```

```

ro<-cor(base$Longitud[1:90],base$Pesos[1:90])
nboot<-500
alpha<-0.05
bas<-cbind(base$Longitud[1:90],base$Pesos[1:90])
mvShapiro.Test(bas)
dat<-cbind(base$Longitud[1:60],base$Pesos[1:60])
datref<-cbind(base$Longitud[61:90],base$Pesos[61:90])
n<-nrow(dat)
m0<-nrow(datref)

bootapl<-function(data,nboot,data.ref=NULL,
                  func=func){
  if(!is.null(data.ref)){
    n<-nrow(datref)
  }
  else{
    n<-nrow(data)
  }
  rs<-matrix(NA,ncol=n,nrow=nboot)
  rss<-numeric(n)
  for (k in 1:nboot)
  {
    if(!is.null(data.ref)){
      dep.ref<-func(data)$dep
      dep<-func(data.ref,data)$dep
      rank <- sapply(dep,function(x) mean(dep.ref <= x))
    }
    else{
      dep.ref<-func(data)$dep
      dep<-func(data)$dep
      rank <- sapply(dep,function(x) mean(dep.ref <= x))
    }
    rs<-tsbootstrap(dep,nb = nboot,type="stationary",b=5)
  }
  return(rs)
}

```

```

}

###Gráfico de control estándar  $T^2$  Hotelling
#FASE I
data.mqcd <- mqcd(dat)
res.mqcs <- mqcs.t2(data.mqcd)
summary(res.mqcs)
plot(res.mqcs, title ="Gráfico  $T^2$  de Hotelling para palitos de chupete")
lc<-res.mqcs$limits[2]
#FASE II
dowl<-mqcd(datref)
res.mqcs2 <- mqcs.t2(dowl)
mean(res.mqcs2$statistics>lc)

###Bootstrap muestra de monitoreo (FASE I)
datn<-dat[-c(41,43,49,50,54,55),]
n<-nrow(datn)
g = c(rep(2, n), rep(3, m0))
btdep<-bootapl(data=dat,nboot = nboot,func=mdepth.SD)
lcboot<-Lim(btdep,alpha = alpha)$LCI
#Gráfico de control profundidad (FASE II)
depth.x1<-mdepth.SD(datn)$dep
depth.x2<-mdepth.SD(datref,dat)$dep
depth.x<-c(depth.x1,depth.x2)
plot(depth.x1, lwd = 0.5, lty = 2, type="b",pch=16,col="blue",
      main="Fase I observaciones correlacionadas",xlab="muestras",ylab="profundidad"
)
abline(h=lcboot,col=4)
out<-(depth.x1<=lcboot)+1
points(x=1:n,depth.x1,col=out)
#Gráfico de control profundidad muestra monitoreo y referencia (FASE II)
plot(depth.x, lwd = 0.5, lty = 2, type="b",col = g,pch=16,
      main="Gráfico control profundidad observaciones correlacionadas",xlab="muestra
)
abline(h=lcboot,col=1)
out<-(depth.x2<=lcboot)+1
points(x=(n+1):(m0+n),depth.x2,col=out)

```

```

#Bootstrap muestra de referencia
btdepref<-bootapl(data=dat,nboot = nboot,data.ref = datref,func=mdepth.SD)
#porcentaje de rechazo profundidad
Lim(btdepref,LCI=lcboot)$RP

#####
#simulando observaciones de una distribución normal
mu1<-mean(base$Longitud)
mu2<-mean(base$Pesos)
var1<-sd(base$Longitud)^2
var2<-sd(base$Pesos)^2
cov1<-cov(base$Longitud,base$Pesos)
cov2<-cov(base$Pesos,base$Longitud)
Sigma<-matrix(c(var1,cov1,cov2,var2),2,2)
mu<-c(mu1,mu2)
#Función simuladora observaciones dependientes
sim.norm<-function( n = 100, p = 2, ro = 0.7, mu = mu,
                    depend=NULL,Sigma=Sigma){
  # Covarianzas
  x = seq(0, 1, length = p)
  x.dist <- as.matrix(dist(x))
  x.cov <- exp(-x.dist)
  # Factorización de la matriz de covarianzas
  U <- chol(x.cov)
  #set.seed(54321)
  z <- t(mvrnorm(n=n,mu=mu,Sigma=Sigma))
  #matrix(rnorm(n * p), nrow = p)
  if(depend){
    error <- t(U) %*% z
    y <- matrix(NA, nrow = p,ncol = n)
    y[,1] <- error[,1]
    for (i in 2:n) {
      y[,i] <- ro*error[,i-1]+error[,i]
    }
    y <- y + mu
  }
}

```

```

else{
  error <- t(U) %*% z
  y<-error+mu
}
data.new <- t(y)
return(data.new)
}

set.seed(123)
#FASE II
prGC<-numeric(5)
nref2<-datref[1:30,]
pal<-mqcd(nref2)
pal.mqcs2 <- mqcs.t2(pal)
for (i in 1:5){
  fas2stat<-pal.mqcs2$statistics+matrix(rep(i,times=30),nrow=30,byrow = TRUE)
  prGC[i]<-mean(fas2stat>lc)
}
prGC
#MEDIDA DE PROFUNDIDAD
#FASE I
xalt<-sim.norm(n=100,mu=mu,depend = TRUE,Sigma = Sigma,ro=ro)
btdep<-bootapl(data=xalt[1:60,],nboot = nboot,func=mdepth.SD)
lcboot<-Lim(btdep,alpha = alpha)$LCI
mualt<-c(mean(xalt[,1]),mean(xalt[,2]))
#FASE II
PR<-numeric(5)
for (i in 1:5) {
  mualt2<-c(mualt[1]+i/2,mualt[2])
  xalt[65:100,]<-sim.norm(n=36,mu=mualt2,depend = TRUE,Sigma = Sigma,ro=ro)
  btdeprefvar<-bootapl(data=xalt,nboot = nboot,data.ref = xalt[61:100,],func=mdepth
#porcentaje de rechazo
  PR[i]<-Lim(btdeprefvar,LCI=lcboot)$RP
}
PR

```

```
PR2<-numeric(5)
for (i in 1:5) {
  mualt2<-c(mualt[1]+i/2,mualt[2]+i)
  xalt[65:100,]<-sim.norm(n=36,mu=mualt2,depend = TRUE,Sigma = Sigma,ro=ro)
  btdeprefvar<-bootapl(data=xalt,nboot = nboot,data.ref = xalt[61:100,],func=mdepth
  #porcentaje de rechazo
  PR2[i]<-Lim(btdeprefvar,LCI=lcboot)$RP
}
PR2
```