



Universidad
Carlos III de Madrid

GRADO EN INGENIERÍA EN TECNOLOGÍAS
INDUSTRIALES: ORGANIZACIÓN INDUSTRIAL

PROYECTO FIN DE GRADO

GRÁFICOS DE CONTROL SIMULTÁNEOS DE
LOCALIZACIÓN Y ESCALA

Autor: Carlos Rodríguez Boada

Tutor: Ignacio Cascos Fernández

Leganés, octubre de 2015

Título: Gráficos de control simultáneos de localización y escala

Autor: Carlos Rodríguez Boada

Director: Ignacio Cascos Fernández

EL TRIBUNAL

Presidente: _____

Vocal:

Secretario:

Realizado el acto de defensa y lectura del Proyecto Fin de Carrera el día __ de _____
de 20__ en Leganés, en la Escuela Politécnica Superior de la Universidad Carlos III de
Madrid, acuerda otorgarle la CALIFICACIÓN de

VOCAL

SECRETARIO

PRESIDENTE

Agradecimientos

Agradezco a mi familia su apoyo durante la finalización de mis estudios y realización del proyecto, con especial mención a Cristina, sin ellos esto no hubiera sido posible.

Quiero agradecer también a mi profesor Ignacio Cascos Fernández por la ayuda prestada, la disponibilidad y la paciencia que ha tenido conmigo a lo largo de los meses. Sin él no hubiera llegado a buen puerto el proyecto.

Resumen

El proyecto que se desarrolla tiene por objetivo la implementación de una serie de gráficos de control basados en el concepto de profundidad en el ámbito estadístico. Lo más relevante del proyecto son los gráficos de control basados en la profundidad del zonoide y los gráficos de control basados en la profundidad de localización y escala. Se desarrollan a través de diferentes funciones implementadas a través del entorno de programación R.

Se realiza una comparativa entre los diferentes gráficos de control propuestos, el gráfico de control basado en la profundidad del zonoide, basado en la profundidad de localización y escala,

Palabras clave: zonoide, localización, escala, gráfico, control, Hotelling.

Índice de contenidos

1. Introducción	1
1.1. Objetivos.....	1
1.2. Metodología de trabajo.....	1
1.3. Estructura	4
2. Introducción. Control estadístico de procesos	6
2.1. Calidad en el ámbito estadístico.....	6
2.2. Niveles de calidad en un proceso	7
2.3. Control estadístico de procesos	9
3. Gráficos de control	13
3.1. Gráficos de control de Shewhart.....	13
3.2. Gráfico de control de Shewhart para rangos y media con datos unidimensionales	16
3.3. Gráfico de control de Shewhart para la media y la desviación típica.....	22
4. Profundidad.....	26
4.1. Profundidad semiespacial	27
4.2. Profundidad simplicial	28
4.3. Profundidad de Mahalanobis	28
4.4. Profundidad del zonoide	29
5. Gráficos de control para la media basados en la profundidad del zonoide	31
5.1. Gráfico basado en la profundidad del zonoide para datos unidimensionales	32
5.2. Estudio del gráfico de control basado en la profundidad del zonoide para datos en dos dimensiones	36
5.3. Estudio del gráfico de control basado en la profundidad del zonoide para datos en tres dimensiones	40
6. Profundidad de localización y escala	43
7. Comparativa para datos unidimensionales	48
8. Comparativa entre profundidad zonoide y T^2 de Hotelling para datos bidimensionales	51
8.1. Comparativa a través de datos procedentes de distribuciones normales	52
8.2. Comparativa a través de datos procedentes de distribuciones exponenciales	54
8.3. Comparativa a través de datos procedentes de distribuciones de Cauchy.....	56
9. Presupuesto.....	60
10. Conclusiones.....	61
11. Bibliografía.....	63

Índice de figuras

Figura 1: "Diagrama de Pareto teórico"	10
Figura 2: "Diagrama de concentración de defectos de una nevera"	11
Figura 3: "Diagrama causa/efecto teórico"	11
Figura 4: "Gráfico de control clásico para rangos"	19
Figura 5: "Gráfico de control clásico para la media"	20
Figura 6: "Gráfico de control clásico para el rango, fase II"	21
Figura 7: "Gráfico de control para la media, fase II"	22
Figura 8: "Gráfico de control clásico para la desviación típica"	24
Figura 9: "Gráfico de control para la media"	25
Figura 10: "Profundidad del zonoide teórica"	30
Figura 11: "Histograma para las muestras de los datos históricos"	32
Figura 12: "Gráfico de control basado en la profundidad del zonoide ($\alpha=0.0027$ y $n=4$)"	34
Figura 13: "Gráfico de control basado en la profundidad del zonoide ($\alpha=0.01$ y $n=2$)"	35
Figura 14: "Gráfico de control basado en la profundidad del zonoide ($\alpha=0.05$ y $n=5$)"	36
Figura 15: "Datos históricos"	37
Figura 16: "Gráfico de control basado en la profundidad del zonoide, datos bidimensionales ($\alpha=0.0027$, $n=4$)"	38
Figura 17: "Gráfico de control basado en la profundidad del zonoide con datos bidimensionales ($\alpha=0.01$, $n=2$)"	39
Figura 18: "Gráfico de control basado en la profundidad del zonoide con datos bidimensionales ($\alpha=0.05$, $n=2$)"	39
Figura 19: "Gráfico de control basado en la profundidad del zonoide con datos tridimensionales ($\alpha=0.0027$, $n=3$)"	41
Figura 20: "Gráfico de control basado en la profundidad del zonoide con datos tridimensionales ($\alpha=0.01$, $n=3$)"	41
Figura 21: "Gráfico de control basado en la profundidad del zonoide con datos tridimensionales ($\alpha=0.05$, $n=4$)"	42
Figura 22: "Gráfico de profundidad de localización y escala (alfa = 0.0027, k = 5)"	45
Figura 23: "Gráfico de profundidad de localización y escala (alfa = 0.01, k = 2)"	46
Figura 24: "profundidad de localización y escala (alfa = 0.05, k = 2)"	47
Figura 25: "Gráfico de control de la media para datos exponenciales, profundidad del zonoide"	49
Figura 26: "Gráfico de la media basado en la profundidad de localización y escala, datos exponenciales"	49
Figura 27: "Gráfico de control basado en la profundidad del zonoide, datos normales"	52
Figura 28: "Gráfico de control basado en el T^2 de Hotelling, datos normales"	53
Figura 29: "Gráfico de control basado en la profundidad del zonoide, datos exponenciales"	54
Figura 30: "Gráfico de control basado en el T^2 de Hotelling, datos exponenciales"	55
Figura 31: "Datos históricos en dos dimensiones"	56
Figura 32: "Gráfico de control basado en la profundidad del zonoide, datos Cauchy"	57
Figura 33: "Gráfico de control de medias basado en el método de Hotelling"	58

Índice de tablas

Tabla 1: "Índice de capacidad y porcentaje de datos fuera de los límites"	8
Tabla 2: "Nº datos fuera de control en función del nivel de calidad"	8
Tabla 3: "Constantes para gráficos de control de Shewhart"	66
Tabla 4: "Constantes para gráficos de control basados en la profundidad del zonoide, datos en 1 dimensión"	67
Tabla 5: "Constantes para gráficos de control basados en la profundidad del zonoide, datos en 2 dimensiones"	67
Tabla 6: "Constantes para gráficos de control basados en la profundidad del zonoide, datos en 3 dimensiones"	68
Tabla 7: "Constantes para los gráficos de control basados en la profundidad de localización y escala" ...	68
Tabla 8: "Funciones creadas en R que aparecen durante el proyecto y definición de sus principales argumentos"	73
Tabla 9: "funciones destacadas en la realización del proyecto"	82



1. Introducción

1.1. Objetivos

El control de calidad en un proceso de producción sirve para detectar posibles anomalías en dicho proceso. A lo largo del presente proyecto se construirán gráficos de control que nos ayudarán a detectar dichas anomalías.

En concreto, el proyecto tiene por objetivo el desarrollo del software necesario para realizar y analizar datos tanto unidimensionales como multidimensionales, a partir de una serie de gráficos de control tales como: el gráfico de control basado en la profundidad del zonoide, el gráfico de control basado en la profundidad de localización y escala, los gráficos basados en el estadístico T^2 de Hotelling y el gráfico de control de Shewhart, ver [1] y [2].

A diferencia de los gráficos clásicos (Shewart), nuestras propuestas se basan en el concepto de profundidad de los datos previstos con respecto a los datos históricos de los que se dispone para el análisis. Mientras que el gráfico de control de la media no usa el concepto de profundidad para analizar los datos a diferencia de los otros.

Frecuentemente se supone que la distribución de los datos de los que se dispone es normal, sin embargo es posible que estos datos provengan de otro tipo de distribuciones, tales como: exponencial, Cauchy, binomial,... Por esto se realizan los diferentes gráficos de control con el fin de evaluar la eficacia de cada uno de éstos frente a los datos procedentes de distintas distribuciones.

Al finalizar el proyecto se mostrará una comparativa entre los diferentes gráficos con el objetivo de mostrar de manera clara, qué tipo de método, dentro de los estudiados, es más adecuado a la hora de analizar cada tipo de datos, ya sean unidimensionales, multidimensionales o procedentes de diferentes distribuciones.

1.2. Metodología de trabajo

Existen técnicas estadísticas que nos permiten estudiar datos con cualquier tipo de distribución e incluso en múltiples dimensiones. Los gráficos de Shewhart solo nos permiten estudiar de manera correcta los datos procedentes de distribuciones normales. Por esto se desarrollan una serie de gráficos de control para poder estudiar no solo datos unidimensionales, procedentes de distribuciones normales, sino para que también se pueden estudiar datos multidimensionales y procedentes de distribuciones no normales.



Se usa el programa informático *R*, software que permite crear gráficos de alta calidad, es gratuito, es actualizado con mucha frecuencia, existe una gran cantidad de información en línea que puede ser consultada y permite realizar una programación clara y sencilla, ver [3].

En la primera parte del proyecto todos los datos usados han sido simulados siguiendo distribuciones normales, ya que los gráficos de control de Shewhart fueron desarrollados para el supuesto de normalidad de los datos como se ha mencionado anteriormente.

En la segunda parte de este proyecto se usarán técnicas válidas para cualquier distribución, con el objetivo de analizar datos multidimensionales, pero manteniendo la hipótesis de la normalidad de estos con el fin de poder comparar datos normales en más de una dimensión.

Los gráficos usados para la representación de datos unidimensionales y multidimensionales se basan en la profundidad de los datos respecto de una medida de centralidad de las muestras, en este proyecto la medida de centralidad usada es la media. Esta profundidad indica la cercanía a la que están las muestras analizadas con respecto a la medida de centralidad usada, cuanto mayor sea esta profundidad (entre 0 y 1) mayor será la cercanía con respecto a este punto central. Se busca que todas las muestras a analizar presenten valores elevados de profundidad con el objetivo de que todas las muestras escogidas de los datos a analizar se encuentren dentro del límite de alarma de estos gráficos de control multivariantes, tales como los basados en la profundidad de localización y escala, en la profundidad del zonoide o en el estadístico T^2 de Hotelling.

Los límites de alarma de cada uno de los gráficos serán dependientes del nivel de significación usado en cada uno de los diferentes análisis, así como, del tamaño muestral elegido.

Si alguna de las muestras queda por debajo del límite de alarma, el proceso se encontrará fuera de control y se tendrá que proceder a eliminar esta muestra o anomalía con el fin de que el proceso se encuentre bajo control.

Todos estos gráficos han sido realizados a través de diferentes paquetes del entorno *R*, tales como el paquete `MSQC` que permite la implementación del gráfico basado en el estadístico T^2 de Hotelling, el paquete `dda1pha`, que permite calcular la profundidad del zonoide y la profundidad de localización y escala. Y el paquete `qcc` que da la posibilidad de realizar gráficos de control clásicos, tales como los de Shewhart, ver [14], [15], [16].



El proyecto finaliza realizando una comparativa entre los diferentes gráficos de control expuestos en él, con el fin de ver cuál de ellos es más eficiente en función del tipo de datos que sean proporcionados para el estudio de los mismos.



1.3. Estructura

División del proyecto a lo largo de los diferentes capítulos.

- **Capítulo 2**

En este capítulo se presentan unas nociones básicas sobre lo que es la calidad y el control de procesos a través de diferentes herramientas usadas para esto.

- **Capítulo 3**

Se presentan los gráficos de control de Shewhart para datos unidimensionales, analizando la media, el rango y la desviación típica de un proceso de producción y sus previsiones.

- **Capítulo 4**

Introduce el concepto de profundidad en el campo estadístico, así como, se introducen diferentes tipos de profundidad, ahondando más en la profundidad del zonoide.

- **Capítulo 5**

Se explican los gráficos de control para la media de un conjunto de datos, tanto unidimensionales como multidimensionales, basados en la profundidad del zonoide.

- **Capítulo 6**

Explica la profundidad de localización y escala. Se realizan diferentes gráficos de control de la media basados en esta profundidad.

- **Capítulo 7**

Se realiza la comparativa entre los gráficos de control para la media de Shewhart, los gráficos para la media basados en la profundidad del zonoide y aquellos basados en la profundidad de localización y escala de datos procedentes de una distribución no normal.

- **Capítulo 8**

Se muestra la comparativa para datos multidimensionales procedentes de diferentes distribuciones no normales a partir de los gráficos basados en la profundidad del zonoide y los gráficos de control de Hotelling.



- **Capítulo 9**

Presupuesto del proyecto incluyendo una estimación del tiempo empleado en su realización.

- **Capítulo 10**

Conclusiones obtenidas al realizar la comparativa entre los diferentes gráficos de control citados en él, además de una visión personal del proyecto.

- **Bibliografía**

Referencias bibliográficas citadas a lo largo del proyecto.

- **Anexos**

En los anexos se muestran diferentes tablas, entre las cuales se encuentran: los valores de las diferentes constantes para los gráficos de control de Shewhart, los valores de éstas para la profundidad del zonoide y para la profundidad de localización y escala, la tabla que refleja la descripción de las diferentes funciones en R usadas a lo largo del proyecto, así como el código de algunas de las más importantes y los datos usados para los gráficos de control de Shewhart y los del zonoide para datos unidimensionales.



2. Introducción. Control estadístico de procesos

Se explicarán brevemente una serie de conceptos básicos para facilitar la lectura del proyecto.

2.1. Calidad en el ámbito estadístico

Para poder definir calidad es necesario conocer que ésta depende de varias dimensiones: desempeño, confiabilidad, durabilidad, facilidad de servicio, estética, características incluidas, calidad percibida y la conformidad con los estándares.

Teóricamente se define calidad como la adecuación de un producto para su uso.

La calidad de conformidad es el grado en que un producto se ajusta a las especificaciones de diseño. Es influido por múltiples factores, como: los procesos de manufactura, el material usado para evaluar la calidad, el índice de calidad total del proceso, etc. Ver [1].

El concepto de calidad es muy importante, ya que una utilización provechosa de ésta lleva a mejorar la estrategia de los negocios que se desarrollaran así como a un crecimiento y a un mejor posicionamiento de la empresa.

Pero esta definición tiene un defecto, no tiene en cuenta la variabilidad de los procesos, por lo que también se puede afirmar que la calidad es una medida inversamente proporcional a la variabilidad de los datos, siendo la variabilidad la medida de la dispersión de las observaciones en torno a la media. Esto quiere decir que cuanta mayor sea la variabilidad que aparece al realizar el proceso, menor será la calidad de éste.

Hablamos de variabilidad porque hay dos tipos de causas de ésta: causas asignables y causas no asignables, ver [19].

Las primeras son causas que pueden ser identificadas, detectadas y, si es posible, eliminadas cuanto antes, con el objeto de que los productos salgan del proceso con los requerimientos impuestos por los clientes, o por los estándares de calidad fijados. Algunas de estas causas pueden ser: temperatura de fusión de un material, velocidad de trabajo, materia prima usada, variabilidad en la dureza del material, etc.

Mientras que las causas no asignables, son aquellas inherentes al proceso mismo, por lo que no pueden ser eliminadas, pero por otro lado, producen un efecto pequeño en la variación total del proceso. Solo podrán eliminarse, modificando el proceso, lo cual normalmente no es económico.



Si la variabilidad del proceso viene de causas comunes de variación, se puede suponer que se trata de una distribución simétrica, mientras que una distribución asimétrica es la que presenta aquellos datos con múltiples ponderaciones por encima o por debajo de la media y son mucho más complicadas de analizar y obtener mejoras.

Normalmente la calidad se evalúa en función de una serie de especificaciones, que en función del producto que queramos estimar, serán unas u otras.

Dentro de esta serie de especificaciones, es necesario definir el valor deseado al que es querido que se ajuste el proceso que se denomina valor nominal, y los valores máximo y mínimo permitidos dentro de las especificaciones del proceso, que son llamados límite superior de la especificación, LSE, y límite inferior, LIE, que dependerá del nivel de calidad elegido.

Mientras que para los demás gráficos de control expuestos en el proyecto estos límites serán nombrados límites de alarma, y el valor al que nos queremos aproximar será también el valor nominal.

Estos límites de alarma serán aquellos que no pueden ser sobrepasados por cada una de las muestras disponibles en el análisis para que el proceso quede bajo control estadístico. Los límites de alarma serán diferentes para cada uno de los gráficos de control que serán realizados, dependen del tamaño muestral escogido además del nivel de significación usado para la realización de estos gráficos.

Si en el conjunto de las muestras del proceso, todas ellas se encuentran dentro de estos límites fijados a través de las ecuaciones, expuestas más adelante, entonces el proceso se encontrará bajo control estadístico. Mientras que si un porcentaje de muestras se encuentra fuera de estos límites, el proceso será inestable, y por tanto se debe analizar qué es lo que ha ocurrido y proceder a eliminar las muestras fuera de control (anomalías) y a realizar de nuevo el control de calidad del mismo.

2.2. Niveles de calidad en un proceso

Como se ha mencionado anteriormente, existen diferentes niveles de calidad para un mismo proceso. Cuanto mayor sea el nivel de calidad, menor número de defectos encontraremos en éste.

Los niveles de calidad se evalúan a través del enfoque "*SIX SIGMA*" [4], que busca disminuir la variabilidad y el número de defectos.



Dentro de este enfoque se tienen diferentes niveles de calidad, desde el 1σ , hasta el nivel máximo de calidad 6σ .

Para conocer cuántos defectos permiten estos niveles, se define la capacidad estadística de los procesos (C_p) como un índice de variabilidad que compara la relación entre el intervalo de especificaciones, LIE y LSE, y una medida de la variación del proceso como es la capacidad natural del proceso. Sólo considera variaciones, sin tener en cuenta si el proceso está o no centrado con respecto a la media.

La manera de obtenerlo matemáticamente es:

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

Conociendo el valor de C_p es posible saber en qué nivel de calidad "SIX SIGMA" está evaluado el proceso.

La tabla siguiente indica el porcentaje de defectos que aparecen en el proceso.

C_p	% datos fuera de la especificación
0,5	13,36 %
0,67	4,55 %
1	0,3 %
1,33	64 ppm
1,67	1 ppm
2	0

TABLA 1: "ÍNDICE DE CAPACIDAD Y PORCENTAJE DE DATOS FUERA DE LOS LÍMITES"

Nivel σ	C_p	Nº de datos fuera de especificación
3	1	66803 por millón
4	1,33	6200 por millón
5	1,67	233 por millón
6	2	3,4 por millón

TABLA 2: "Nº DATOS FUERA DE CONTROL EN FUNCIÓN DEL NIVEL DE CALIDAD"



No se incluye en la tabla anterior los niveles 1σ y 2σ ya que no suelen ser usados debido al alto número de defectos que aparecen.

Como podemos ver en la Tabla 2, el óptimo nivel de calidad buscado es el 6σ , ya que si éste se cumpliera, el proceso sería prácticamente perfecto, porque solo 3 datos de cada millón quedarían fuera de las especificaciones. Este nivel es teórico ya que es imposible de conseguir estándares de calidad tan altos a nivel práctico.

2.3. Control estadístico de procesos

Algunas de las herramientas usadas para el control estadístico de procesos fueron explicadas y puestas en marcha por el profesor japonés Kaoru Ishikawa y el ingeniero estadounidense W. A. Shewhart, ver [1] y [2].

Debe usarse de manera rutinaria con el fin de identificar las oportunidades de mejorar y eliminar el desperdicio que aparezca en el proceso.

El control estadístico de procesos trata de realizar un análisis de la información suministrada por un proceso y detectar las causas de los posibles defectos a través de una serie de herramientas, que serán definidas más adelante.

La hoja de verificación

Fue diseñada con el objeto de facilitar la lectura del resumen de los datos históricos de los defectos producidos a lo largo de un proceso. Es muy útil a la hora de buscar tendencias en los defectos, por ejemplo si la mayoría de los defectos ocurren durante una estación del año, esta hoja lo muestra y puede indicar, por ejemplo, que se debe estudiar a los trabajadores empleados durante dicho periodo de tiempo con el objeto de mejorar su rendimiento o vislumbrar cuales son los fallos de estos durante este periodo.

Es muy importante clasificar los datos que van a analizarse, dando la máxima información de estos, tales como el número de operación, la fecha, el operario, el número de máquinas a utilizar y una serie de datos adicionales.

El diagrama de Pareto

Muestra una distribución de frecuencia en el que los datos están agrupados por categorías (por ejemplo: oxidación de la herramienta, desgaste de la herramienta, tiempo de uso de la herramienta, etc.) y los porcentajes agregados de cada barra se conectan mediante una línea para mostrar la suma incremental de cada categoría respecto del total.

Constituye un sencillo gráfico de análisis que permite discriminar entre las causas que aparecen con más frecuencia y las que aparecen con poca frecuencia.



Suele ser usado para determinar las causas raíz de un problema y decidir el objetivo de mejora y los elementos a mejorar (en el caso de una fábrica: las mayores pérdidas energéticas o los mayores costos de materias primas, por ejemplo).

Debe ser acompañado de *esquemas de ponderación para defectos* con el fin de evaluar los defectos más graves y los menos graves, ya que simplemente con la frecuencia no es posible ver esto.

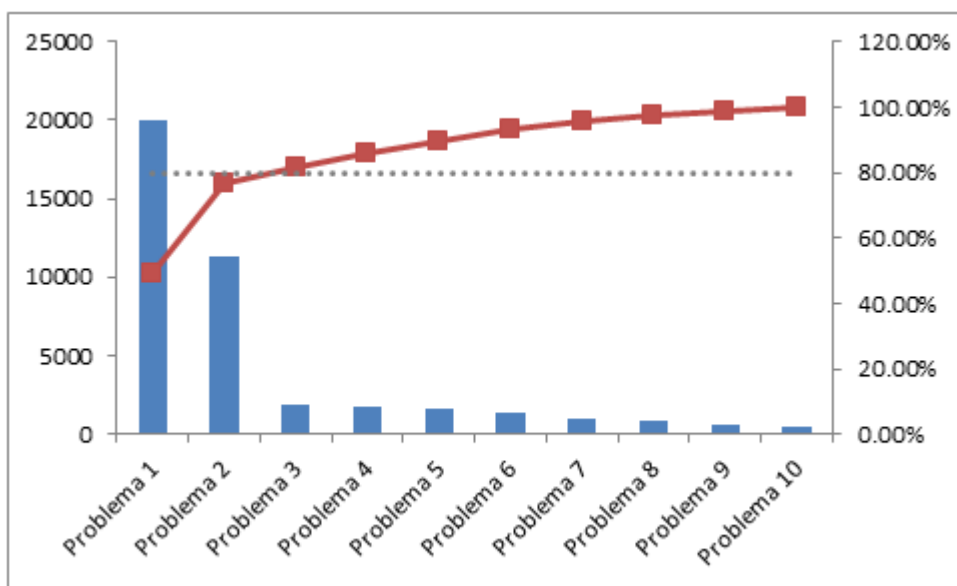


FIGURA 1: “DIAGRAMA DE PARETO TEÓRICO”

Histogramas

Son representaciones gráficas de una variable en forma de rectángulos donde el área de cada rectángulo es proporcional a la frecuencia de los valores en el intervalo representado por la base del rectángulo.

Permite conocer la forma de distribución de los datos, la tendencia del proceso y la dispersión de éste. A través del histograma se obtiene una idea de la posible distribución de los datos que han sido analizados.

En la Figura 11 del presente proyecto se muestra un histograma para unos datos simulados, según una distribución normal.

Diagrama de concentración de defectos

Se usa para determinar si el lugar donde se encuentran los defectos transmite alguna información sobre la causa de éstos. Se representa la pieza en todas las vistas necesarias así como los defectos que existen en ella.



Son usados a nivel industrial sobre todo en actividades de pintura, cizallado y ensamblaje de componentes.

En la siguiente figura, extraída del libro: *"Control Estadístico de la calidad, 3ª ed."* [1] de D.C. Montgomery, se muestran los defectos que aparecen en la superficie de una nevera con el objetivo de evaluar si se deben a un fallo ocasional o si se repite en varios casos.

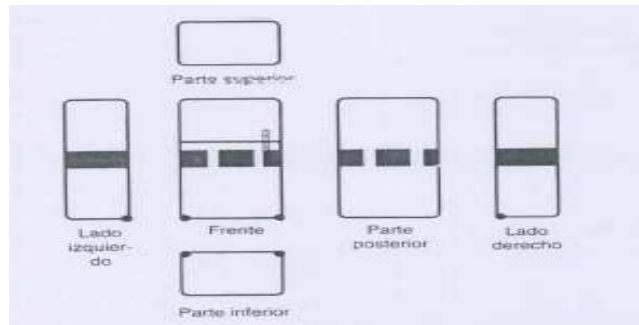


FIGURA 2: "DIAGRAMA DE CONCENTRACIÓN DE DEFECTOS DE UNA NEVERA"

El diagrama de causa y efecto

Es una herramienta que se usa con fin de identificar las posibles causas potenciales de aparición de defectos.

Para crearlo se debe definir el problema a estudiar, analizarlo, colocar el rectángulo principal (efecto) y la línea central, definir las categorías de las causas y colocarlas enlazadas con la línea central, colocar las causas posibles en las categorías definidas anteriormente, clasificar las causas que pueden incidir en el problema y poner en marcha una actitud correctiva.

Se muestra una representación teórica de este diagrama.

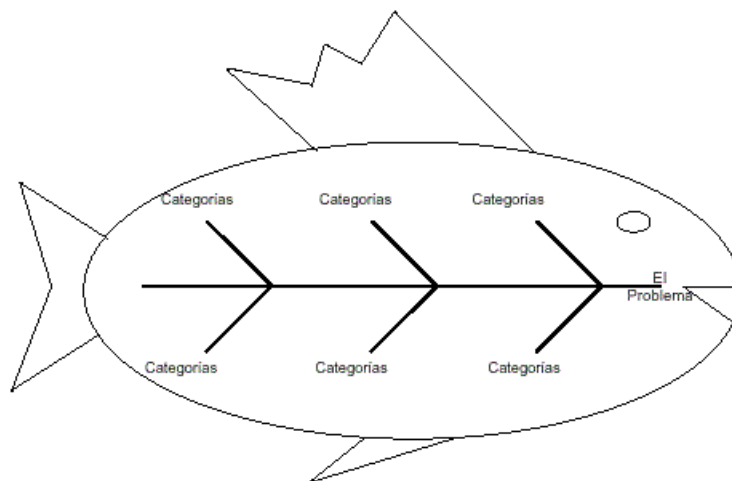


FIGURA 3: "DIAGRAMA CAUSA/EFEECTO TEÓRICO"



Son muy útiles a la hora de trabajar en equipo ya que compromete a los trabajadores a solucionar todos los problemas operativos de manera conjunta consiguiendo una actitud de cooperación por parte de los empleados logrando un incremento en el cumplimiento de los objetivos y en la calidad del proyecto.

Diagrama de dispersión

Es una representación gráfica para datos bidimensionales en un sistema de ejes cartesianos x-y donde cada par x e y se representa por un punto y define una característica a estudiar de nuestro problema.

Se suelen utilizar experimentos específicos para garantizar analizar la relación entre las variables de los datos, por ejemplo: el gráfico puede mostrar una correlación positiva entre dos variables, pero también puede estar influido por otra variable externa al proceso que en ese caso no se está analizando, pero que repercute en la medición de la causalidad.

En la Figura 10 del presente proyecto se muestra un diagrama de dispersión basado en la profundidad del zonoide para un conjunto de 2000 datos.

Gráfico de control

Es un diagrama en el que aparece una línea central o valor nominal y unos límites de control específicos usados para medir la calidad e indican si el proceso está bajo control o fuera de control.

El proceso estará bajo control, siempre y cuando todas las muestras se encuentren dentro de estos límites.

Con los gráficos de control realizaremos el control estadístico de procesos, ya que es una herramienta muy útil y aporta mucha información.

Los primeros gráficos de control fueron propuestos por W. A. Shewhart, que se expondrán en el próximo capítulo, como los que aparecen en las Figuras 4, 5, 6, etc.



3. Gráficos de control

El gráfico de control muestra los valores de la característica de estudio de los datos proporcionados por el proceso a través de un gráfico, con el fin de analizar estos y llegar a detectar causas de posibles defectos; es decir, mejorar la calidad del proceso.

El gráfico de control es representado por un diagrama que muestra el valor medio o línea central de la característica de la calidad que se quiera evaluar y dos límites de control, inferior y superior, que informan si los datos recogidos del proceso se encuentran bajo control o no, esto es, cuando algunos datos del proceso se encuentren fuera de estos límites, el proceso se encontrará fuera de control.

Muestran una información de diagnóstico que permite la consecución de un cambio en el proceso industrial que mejore éste, permite obtener información sobre la capacidad del proceso, son útiles para prevenir defectos y ayuda a mantener el proceso bajo control ahorrando costes de fabricación y costes en el control de calidad.

Al hablar de diseño es muy importante la elección del tamaño de la muestra, los límites de control, la frecuencia de muestreo y el tipo de variabilidad que muestra el proceso, ya que con estos datos se estima si un gráfico de control presenta datos estacionarios o no estacionarios y correlacionados o no.

Para poder aplicar el control estadístico de procesos, éste debe tener una cierta estabilidad que permita obtener un cierto grado de predicción, pudiendo conseguir una serie de mejoras con el fin de desarrollar medidas para el aumento de la calidad.

3.1. Gráficos de control de Shewhart

Los gráficos de control son una de las herramientas más usadas para el control estadístico de procesos. Permiten detectar las causas asignables de anomalías en un proceso y, con ayuda de la ingeniería de proceso, tratar de eliminarlas consiguiendo una mejora del proceso y de la calidad de éste.

Los gráficos de Shewhart son muy útiles cuando los datos de partida son estacionarios y no correlacionados

La línea central indica el promedio de nuestros datos, mientras que los límites de control permiten observar si las muestras de datos están bajo control o no.



Para estudiar esto, diremos que un proceso estará bajo control cuando presente todas las muestras dentro de los límites de control.

Es necesario definir el tamaño de cada muestra y la frecuencia de muestreo ya que cuanto más grande sea la muestra tomada más fácil será la detección de pequeños fallos en el proceso. Así mismo, con la frecuencia ocurre lo mismo, se busca tomar muestras con mucha frecuencia con el fin de detectar las anomalías de forma rápida y precisa.

En la actualidad suelen tomarse muchas muestras de tamaño pequeño pero de manera continuada, con lo que conseguimos monitorizar el control del proceso casi a tiempo real.

Hay dos tipos de errores que pueden aparecer cuando se realizan gráficos de control, no detectar una anomalía cuando ésta está presente o detectarla cuando no existe.

Utilizando la jerga de la inferencia estadística podemos hablar del error tipo *I* y del error tipo *II*, el error de tipo *I* se da cuando una muestra se localiza fuera de los límites de control sin que aparezca ninguna causa asignable, mientras que el error de tipo *II* aparece cuando ésta se localiza dentro de los límites de control pero el proceso está fuera de control. Los errores de tipo *II* son mucho más difíciles de apreciar a simple vista en nuestros gráficos de control.

La hipótesis nula sería que no apareciera ninguna anomalía, mientras que la hipótesis alternativa sería que se presentara cualquier anomalía.

Teóricamente, el error de tipo *I* aparece cuando la hipótesis nula (H_0) es verdadera pero se rechaza, por lo general, fijamos la probabilidad de que este error ocurra como α (nivel de significación). Mientras que el error de tipo *II* ocurre cuando no se rechaza la hipótesis nula (H_0) sabiendo que esta es falsa en el conjunto de los datos, siendo su probabilidad β .

El valor del error de tipo *I* se fija en un nivel bastante pequeño para asegurar una buena calidad.

$P = (\text{escoger } H_1 H_0) = \alpha$	\longrightarrow	Error tipo I
$P = (\text{escoger } H_0 H_1) = \beta$	\longrightarrow	Error tipo II

Para la realización de los gráficos de control de Shewhart es necesario conocer si los datos a estudiar pertenecen a una distribución normal, que se demostrará a través de un test de bondad de ajuste, ver [6], [7].



Siempre y cuando el test de bondad sea positivo ($p\text{-valor} \geq 0.05$) y se fije el nivel de significación α , es posible calcular los límites de control, LSC y LIC, para decidir si estos datos se encuentran bajo control o fuera de control, siguiendo lo anteriormente expuesto.

Reglas de Shewhart para el control del proceso

Para saber si los datos están bajo control existen una serie de reglas, algunas de ellas definidas por la empresa Western Electric, ver [5]:

- Uno o más puntos quedan fuera de los límites de control.
- Catorce puntos seguidos alternándose arriba y abajo.
- Seis puntos seguidos que se incrementan o decrecientan de manera sostenida.
- Uno o más puntos cerca de un límite de control.
- Ocho puntos consecutivos se localizan en un mismo lado de la línea central.

Cada uno de los puntos que se mencionan en las reglas, son cada una de las muestras tomadas de los datos del proceso.

Todos estos criterios y los anteriormente citados son usados para descubrir las muestras que estarán fuera de control, y por tanto proceder a su retirada, con el objeto de conseguir la estabilidad del proceso.

Aproximación teórica

Se realiza una primera aproximación teórica a los gráficos de control de Shewhart.

Para ello es necesario suponer un estadístico, que será llamado w y representa una característica del proceso a estudio.

μ_w es la media de la característica y σ_w es la desviación típica.

Se define la línea central, y los límites de control del proceso para esta característica. En este caso la línea central es la media de las muestras del proceso.

$\begin{aligned} \text{Línea central} &= \mu_w \\ \text{Límite superior de control (LSC)} &= \mu_w + L * \sigma_w \\ \text{Límite inferior de control (LIC)} &= \mu_w - L * \sigma_w \end{aligned}$

Siendo L la distancia entre la media y los límites de control teóricos.

Se realizan una serie de gráficos de control clásicos de Shewhart para exponer lo escrito de manera más clara.



3.2. Gráfico de control de Shewhart para rangos y media con datos unidimensionales

Este tipo de gráfico monitoriza la media muestral de un número de muestras m de tamaño n de la característica a estudiar en el proceso, ver [17], [19].

El gráfico de control puede comenzar a realizarse si los resultados obtenidos con el test siguen las características de una distribución normal. Si no es así, deben transformarse los datos, en busca de normalidad aplicando el teorema del límite central.

Para comprobar la normalidad se debe realizar el test de bondad de ajuste. La hipótesis nula (H_0) del test de bondad es que todos el conjunto de datos se aproximen a una distribución normal, mientras que la hipótesis alternativa (H_1) será que los datos no se aproximen a una distribución normal, sino a otro tipo de distribución.

Siempre y cuando el test de bondad nos devuelva un valor mayor que el nivel de significación α que hemos escogido, la hipótesis nula (todos los datos se ajustan a una normal) no se rechaza y se podría comenzar a graficar.

Este test se realiza a través de la función `shapiro.test(x)`, incluida en el programa estadístico R, siendo x el conjunto de datos a estudiar de nuestro proceso. Esta función devuelve un p -valor que debe ser mayor del nivel de significación escogido con el fin de poder admitir que los datos a estudio provienen de una distribución normal, y por tanto para poder empezar a realizar el gráfico.

Si conocemos la media y la desviación típica de la característica a estudiar podemos calcular la desviación estándar de ésta y realizar el gráfico con unos límites de control y una línea central, que en este caso serán el rango y la media del proceso.

Conociendo la media y la desviación típica de una variable x , que sigue una distribución normal $N(\mu, \sigma)$, podemos calcular la desviación estándar de nuestra variable como: $\frac{\sigma}{\sqrt{n}}$ siendo n el número total de datos del proceso.

Con esto es posible calcular los límites de control y el límite central, y proceder a analizar si nuestro proceso se encuentra o no bajo control. Estos son los límites de control para la media del proceso (μ).



$$\text{Límite de control superior (LSC, en adelante)} = \mu + z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

$$\text{Línea central (LC, en adelante)} = \mu$$

$$\text{Límite de control inferior (LIC, en adelante)} = \mu - z_{\alpha/2} * \frac{\sigma}{\sqrt{n}}$$

Para un nivel de significación $\alpha = 0.0027$, el valor de $z_{\alpha/2} = 3$, ya que se busca un muy alto nivel de calidad. Para diferentes valores del nivel de significación escogido, el valor de $z_{\alpha/2}$ variará, siendo el máximo valor el obtenido para un nivel $\alpha = 0.0027$.

Se debe comenzar decidiendo si todos los datos del proceso se ajustan a una normal a través de un test de bondad, como se ha mencionado anteriormente.

Para el estudio se dispone de m muestras de tamaño n . Una vez admitido que siguen una distribución normal, se calcula la media de cada una de las muestras con el fin de analizar la característica en el proceso, en este caso se analiza la media y el rango.

$$\bar{x}_i = \frac{x_{i1} + x_{i2} + \dots + x_{in}}{n}$$

También se calcula el rango de las muestras y el promedio de los rangos que será usado para estimar σ .

El rango es la diferencia entre el valor mayor y menor de cada muestra. Cada una de las muestras tendrá un rango determinado. Se calcula la media de todos los rangos.

$$R_n = \text{Valor máximo} - \text{Valor mínimo}$$

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

Promedio de las medias de las muestras ($\bar{\bar{x}}$), con el fin de analizar el proceso a través del promedio de las muestras.

$$\bar{\bar{x}} = \frac{\bar{x}_1 + \bar{x}_2 + \dots + \bar{x}_m}{m}$$

Límites de control tanto para el gráfico de la media como para el gráfico de control de rangos.



Rangos

$$\begin{aligned} \text{LSC} &= \bar{R} * D_4 \\ \text{LSI} &= \bar{R} * D_3 \\ \text{LC} &= \bar{R} \end{aligned}$$

Medias

$$\begin{aligned} \text{LSC} &= \bar{x} + A_2R \\ \text{LC} &= \bar{x} \\ \text{LIC} &= \bar{x} - A_2R \end{aligned}$$

Los coeficientes A_2, D_4, D_3, d_3, d_4 vienen tabulados en la tabla 3 que se adjunta en el anexo I para múltiples valores de n, siendo:

$$\begin{aligned} D_4 &= 1 + z_{\alpha/2} \left(\frac{d_3}{d_2} \right) \\ D_3 &= 1 - z_{\alpha/2} \left(\frac{d_3}{d_2} \right) \end{aligned}$$

Se fija el valor de significación $\alpha = 0.0027$, con lo que el valor de $z_{\alpha/2} = 3$ y se puede comenzar a calcular los límites de control del proceso. Estos valores pueden ser modificados con el objetivo de estudiar otros niveles de calidad.

Cuando los resultados de los análisis de la muestra de control a lo largo del tiempo se encuentran dentro de los límites aceptados, el sistema se encuentra bajo control estadístico. Pero cuando se encuentran muestras fuera de los límites especificados, o se encuentran tendencias o patrones en los datos el sistema se encuentra fuera de control.

Ejemplo práctico:

El ejemplo se ha realizado suponiendo una distribución normal de datos unidimensionales con media 35 y desviación típica 0.5.

Los datos deben ser unidimensionales, ya que este tipo de gráfico solo nos da cierta fiabilidad en el estudio para datos unidimensionales.

Para el ejemplo, se busca conocer si una serie de muestras previstas estarán dentro de las especificaciones de diseño. En la primera parte, con los datos simulados buscamos unos límites de control aceptables, con el objetivo de analizar las muestras previstas.



En la segunda parte, una vez que hemos monitorizado el proceso, y conseguido unos límites de control aceptables, se evalúan las previsiones de datos que se tienen con el objetivo de decidir si el proceso futuro puede estar bajo control.

Fase 1.

En primer lugar se ha de comprobar que los datos simulados se ajustan a una distribución normal. Para ello se realiza el test de Shaphiro-Wilk, a través de la función del programa R: `shapiro.test (datos1)`.

```

Shapiro-Wilk normality test

data:  datos1
W = 0.9869, p-value = 0.2769
    
```

Como el “*p-valor*” de los datos es mayor que el nivel de significación α escogido, se acepta la hipótesis nula (se ajustan a una normal) y se realiza el gráfico a través de la función: `graficorangos (datos1, 2, 0.0027)`.

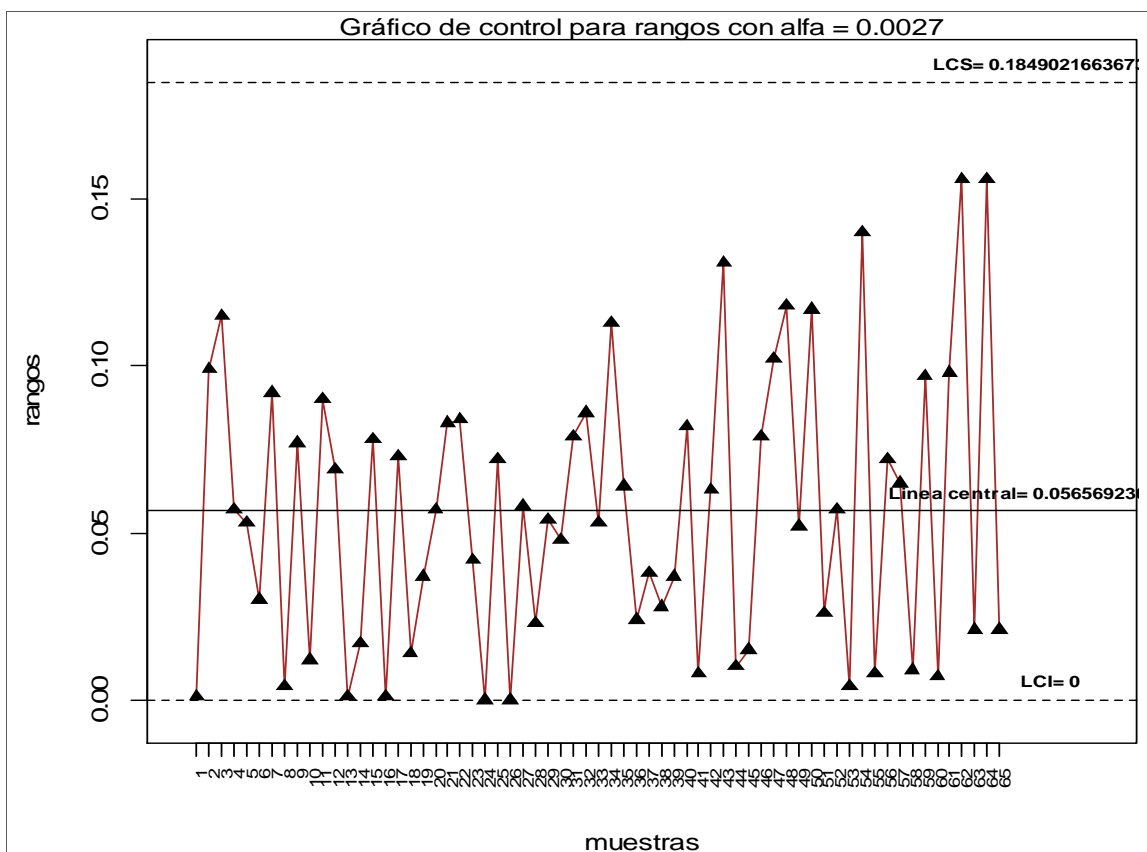


FIGURA 4: “GRÁFICO DE CONTROL CLÁSICO PARA RANGOS”

Función en R: `graficorangos (datos1, 2, 0.0027)`



Como se ve en el gráfico, cogiendo un nivel de significación de 0.0027, ninguno de los datos sobrepasa los límites de control. Se afirma que no existe variabilidad y por tanto se toman como aceptables estos límites de control, asumiendo la estabilidad del proceso para el caso del rango.

Una vez realizado el gráfico de control de rangos, se realiza el gráfico de control del proceso para la media, a través de una función creada en R para tal fin: `graficomedias (datos, n, alfa)`.

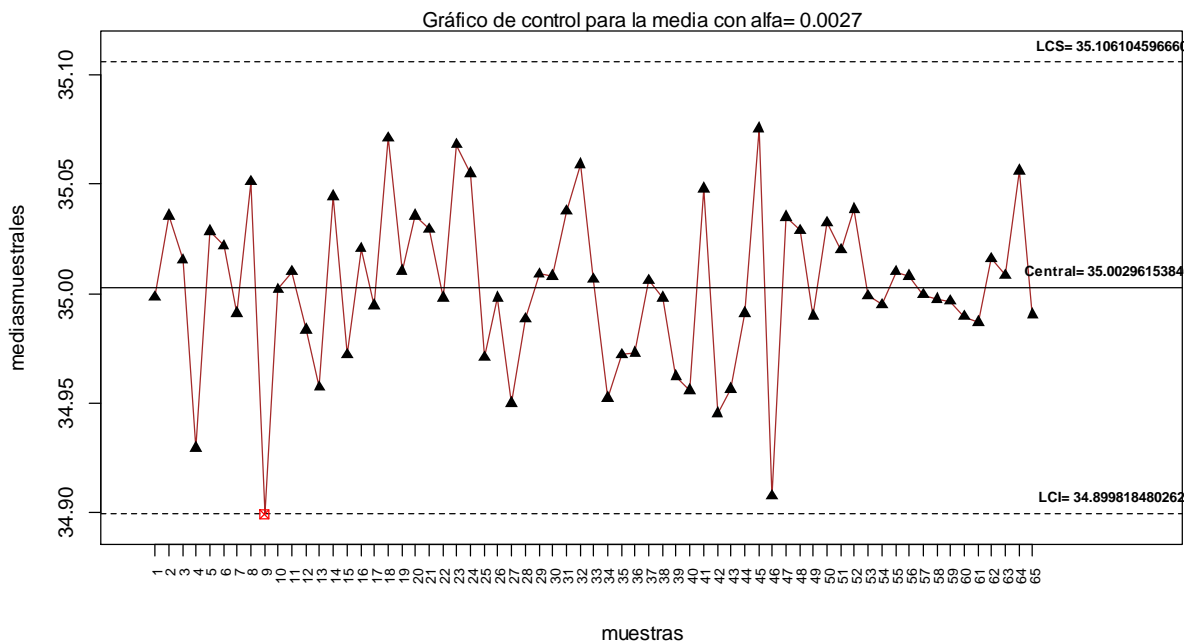


FIGURA 5: “GRÁFICO DE CONTROL CLÁSICO PARA LA MEDIA”

Función en R: `graficomedias (datos1, 2, 0.0027)`

Para el caso del gráfico de la media existe una muestra fuera de control, la muestra 9, se debe eliminar esta anomalía y volver a realizar el control del proceso para obtener unos límites de control aceptables.

Después de calcular los nuevos límites de control para la media con un total de 64 muestras, ya que la muestra 9 ha sido eliminada, son:

$$LSC = 35.10707021018$$

$$LC = 35.00457812$$

$$LIC = 34.902086039818$$



Mientras que los nuevos límites de control para el rango son:

$$LSC = 0.178138679455047$$

$$LC = 0.0545$$

$$LIC = 0$$

Se puede comenzar con la fase II ya que se dispone de unos límites de control aceptables, después de eliminar la muestra fuera de control.

Fase 2

Sabiendo esto, y habiendo obtenido unos límites de control aceptables, se estudian una serie de muestras de datos previstas que aparecen en el Anexo IV.

Se realizan los gráficos de la media y del rango con las mismas funciones que anteriormente, pero colocando en ésta los límites de control calculados como válidos en la fase anterior.

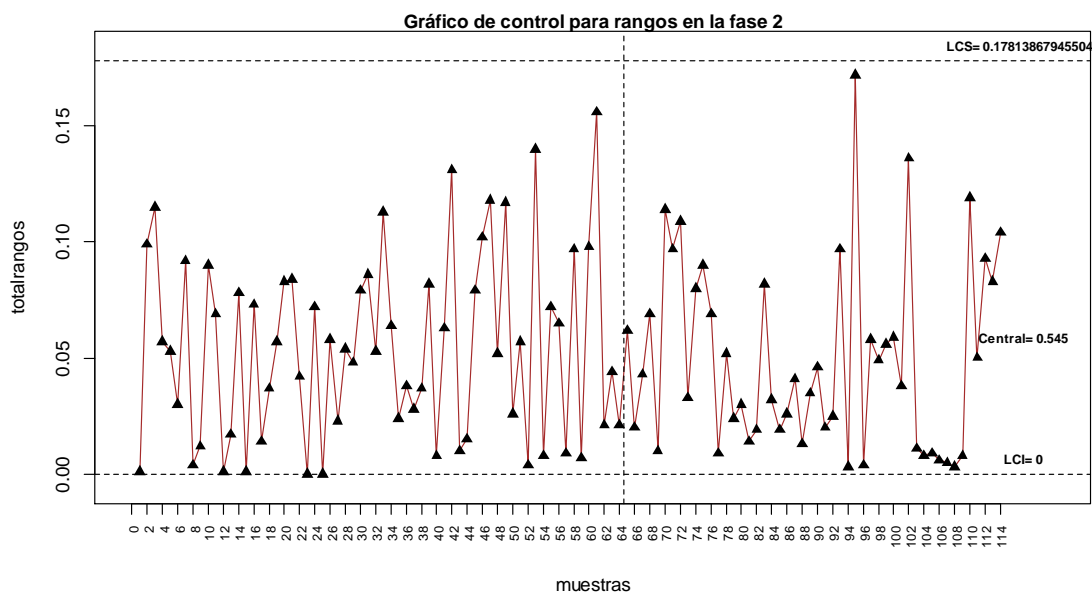


FIGURA 6: “GRÁFICO DE CONTROL CLÁSICO PARA EL RANGO, FASE II”

Función en R: `controlrangos2 (datostotales, 2, 0.178138679455047, 0.545, 0)`

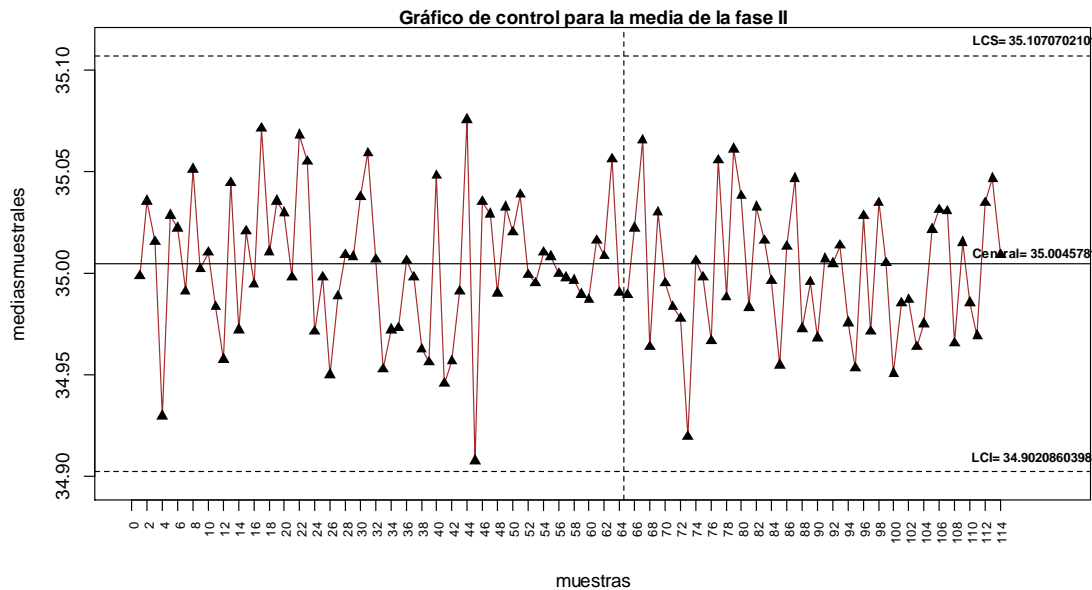


FIGURA 7: “GRÁFICO DE CONTROL PARA LA MEDIA, FASE II”

Función en R: `controlmedias2 (datostotales, 2, 35.10707021018, 35.00457812, 34.902086039818)`

Realizados los gráficos de control para esta segunda fase pasamos a analizar si este proceso se encuentra estable o no.

Se observa que el gráfico de control para rangos de esta segunda fase es estable, ya que todas las muestras escogidas se encuentran dentro de los límites de control, establecidos como válidos en la primera fase.

Lo mismo ocurre en el gráfico de control de la media de esta segunda fase, en el que todas las muestras se encuentran dentro de los límites fijados en la fase anterior por lo que se afirma sin temor a equivocarse que los datos futuros van a cumplir las especificaciones de diseño, y que el proceso se encuentra bajo control.

3.3. Gráfico de control de Shewhart para la media y la desviación típica

El gráfico basado en la desviación típica es usado cuando el tamaño de cada muestra es relativamente grande, o cuando el tamaño de cada una de las muestras va variando. Se usa para detectar pequeños cambios en nuestro proceso.

Es parecido al anterior, pero no se calcula la desviación típica a través de los rangos, sino que se estima directamente.



En cuanto al proceso que hay que llevar a cabo para la realización del gráfico es: conocer si los datos usados para el ejemplo se ajustan a una normal, calcular el promedio muestral y la desviación típica muestral, después calcular los límites de control y por último realizar el gráfico.

Sabiendo que la desviación típica muestral se calcula como:

$$S = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

Denotamos su promedio como: $\bar{S} = \frac{S_1 + S_2 + \dots + S_m}{m}$, ya es posible calcular los límites de control de los dos gráficos.

Límites para el gráfico de control del promedio muestral:

$$\begin{aligned} \text{LSC} &= \bar{\bar{x}} + \frac{3\bar{S}}{C_4\sqrt{n}} = \bar{\bar{x}} + A_3 * \bar{S} \\ \text{LIC} &= \bar{\bar{x}} - \frac{3\bar{S}}{C_4\sqrt{n}} = \bar{\bar{x}} - A_3 * \bar{S} \\ \text{Línea central} &= \bar{\bar{x}} \end{aligned}$$

Límites para el gráfico de control de la desviación típica muestral.

$$\begin{aligned} \text{LCS} &= \bar{S} + 3 \frac{\bar{S}}{C_4} * \sqrt{1 - C_4^2} = B_4 * \bar{S} \\ \text{LIS} &= \bar{S} - 3 \frac{\bar{S}}{C_4} * \sqrt{1 - C_4^2} = B_3 * \bar{S} \\ \text{Línea central} &= \bar{S} \end{aligned}$$

Se calculan los gráficos tanto de la media como de la desviación típica muestral y a partir de estos se indica si los procesos se encuentran o no bajo control, además de evaluar si aparecen causas asignables de defectos, con el fin de controlarlas y si es posible eliminarlas.

Ejemplo práctico:

Los datos escogidos son los mismos que en el ejemplo de la subsección 3.2, y como anteriormente se ha visto se ajustan a una normal, por lo que no es necesario pasarles el test de *Shapiro-Wilk*.

Para este proceso se escoge un nivel de significación $\alpha = 0.0027$ y un tamaño muestral de 2; es decir, 64 muestras de datos unidimensionales.



En el primer gráfico que se muestra, la línea central, llamada *central* en la función creada para tal fin en el programa R, representa el valor medio de la desviación estándar del conjunto de muestras escogidas de los datos usados en el análisis.

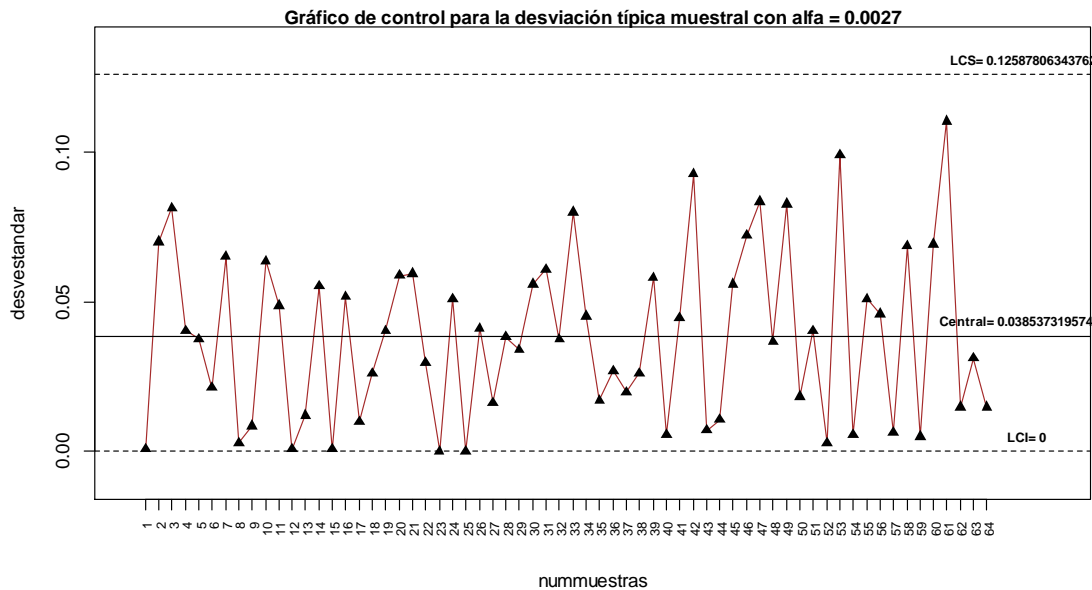


FIGURA 8: “GRAFICO DE CONTROL CLÁSICO PARA LA DESVIACIÓN TÍPICA”

Función en R: `controldesvestandar (datos1, 2, 0.0027)`

A la vista del gráfico se afirma que el proceso es estable porque todas las muestras se encuentran dentro de los límites de control, por lo que la variabilidad del proceso está bajo control para el caso del rango.

Para el cálculo de los límites de control de la media del proceso es necesario el valor del coeficiente c_4 , que viene tabulado, para diferentes valores del tamaño muestral elegido en la tabla 3 que se muestra en el anexo I.

Se debe calcular la media de cada una de las muestras, así como el promedio de las medias muestrales, y una vez hecho esto y conociendo el valor del coeficiente c_4 se calculan los valores de los límites de control tanto superior como inferior de este proceso, y se procede a evaluar la estabilidad del proceso.

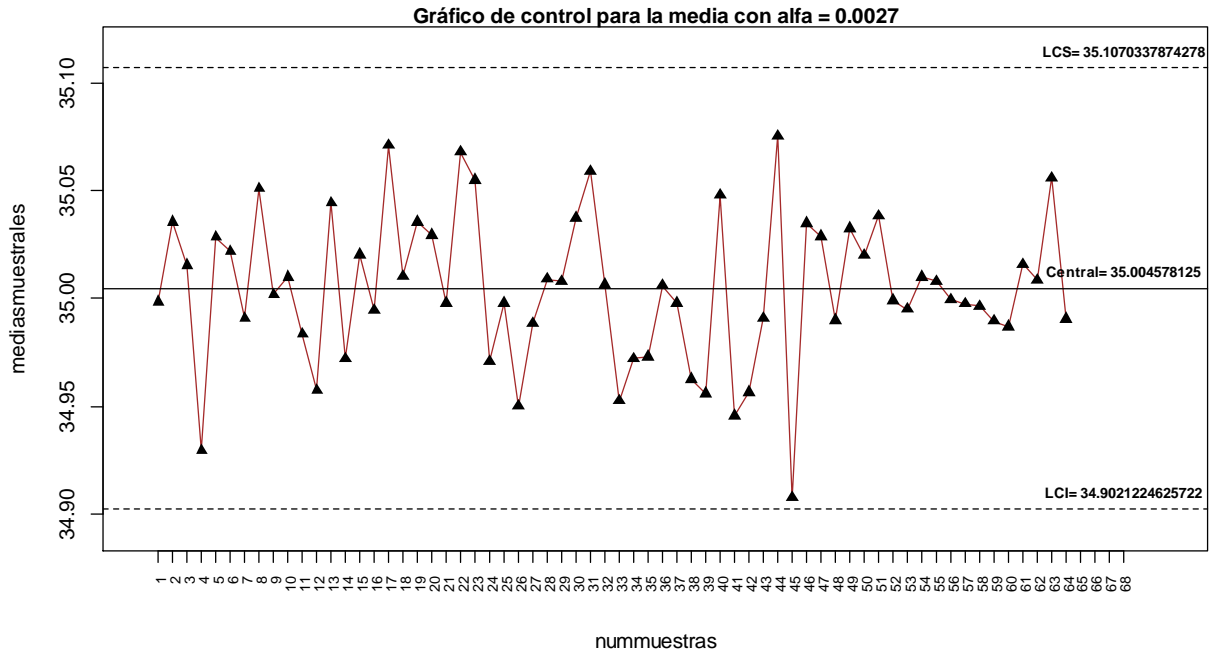


FIGURA 9: “GRÁFICO DE CONTROL PARA LA MEDIA”

Función en R: `controlmedia (datos1, 2, 0.0027)`

En el gráfico de control de la media se encuentran todas las muestras dentro de los límites de control, por lo que a la vista de los dos gráficos se ve que el proceso está bajo control para los niveles establecidos, y se usarán estos límites de control obtenidos para el caso del análisis de datos de previsiones futuras, que no se ejemplifica por ser lo mismo que en la subsección 3.2 pero con datos tridimensionales.



4. Profundidad

La profundidad en el campo de la estadística multivariante se refiere al grado de centralidad de un punto con respecto a una nube de puntos con la misma dimensión o de la distribución de probabilidad de los mismos, ver [25].

El punto más profundo de la nube de puntos escogida es alguna medida de tendencia central, normalmente la media o alguna mediana para el análisis multivariante. La profundidad de un punto o de una muestra disminuye a medida que se aleja del punto más profundo con mayor o menor velocidad según la dirección que se aleje.

En el caso de estudio de datos multivariantes, se escogerá como medida de centralidad un conjunto de puntos que representen la característica a estudiar que será el punto más profundo de la nube. Se calculará la profundidad de estos datos a partir de esa medida de centralidad, ver [23].

Matemáticamente, una función de profundidad es una función acotada $D(\cdot; P) : \mathbb{R}^d \rightarrow [0,1]$, que fija a cada punto de \mathbb{R}^d su grado de centralidad respecto de una distribución de probabilidad P , siendo d la dimensión del espacio de referencia, ver [22].

Los valores con profundidades más grandes, próximas a 1, corresponden al centro de la distribución, mientras que las más pequeñas corresponden a regiones externas, siendo estos los que más posibilidades tienen de estar fuera de control y fragmentar la estabilidad del proceso.

Para que una función de profundidad sea considerada como tal, debe cumplir una serie de propiedades, ver [11] y [25].

1. *Invarianza afín*: La profundidad de un punto respecto de la nube de puntos no depende del sistema de coordenadas elegido.

2. *La función de profundidad se anula en el infinito*: $\lim_{x \rightarrow \infty} D(x; P) = 0$, la profundidad de un punto lo suficientemente lejos del centro de la distribución es 0.

3. *Semicontinuidad superior*: El conjunto de puntos, cuya profundidad es al menos un valor dado, es topológicamente cerrado.

$$\{x \in \mathbb{R}^d : D(x; P) \geq \alpha\} \text{ es cerrado para cualquier } \alpha.$$

4. *Monotonía respecto del punto más central*: Si θ es un punto de profundidad máxima, entonces $D(x; P) \leq D(\theta + \lambda(x - \theta); P)$ para cualquier $0 \leq \lambda \leq 1$, es decir, la profundidad decrece a lo largo de las semirrectas que parten del punto más profundo.



Al hablar sobre la profundidad, se ha mencionado varias veces el *centro* de la distribución a estudio. El valor de la profundidad de éste es 1, y suele ser el promedio de la característica que se quiere estudiar, o un valor al que queramos aproximarnos. Pero, una nube no siempre tiene un centro claramente definido, solo lo tendrá cuando la nube sea simétrica en algún sentido, y ese centro es el punto de simetría.

La profundidad indica la región central más pequeña a la que pertenece x , siendo x la muestra escogida.

Las construcciones que serán llevadas a cabo a través de este y los siguientes capítulos muestran los gráficos de control utilizando el concepto de profundidad, y son válidos para datos en cualquier dimensión y procedentes de cualquier distribución; es decir, no hace falta que éstos provengan de distribuciones normales, como en el capítulo anterior.

Se explican a continuación algunas funciones de profundidad usadas a lo largo de los años, incluyendo la profundidad del zonoide, que será explicada de manera más detallada en los capítulos siguientes.

4.1. Profundidad semiespacial

Tres fueron los autores que hablaron de esta profundidad, Tukey, Rousseeauw y Ruts, ver [8].

$HD(x; P) = \text{ínfimo} \{P(H) : x \in H \text{ semiespacio cerrado}\}$, Siendo H un semiespacio cerrado y x un punto de su interior.

Si en vez de una distribución se tiene una muestra $x_1, x_2, x_3, \dots, x_n$ debe considerarse la distribución empírica respecto de la misma. De esta manera, la profundidad de x respecto de la muestra será la menor fracción de puntos de la muestra contenidos en un semiespacio con x en su frontera.

De manera más gráfica, en el caso bidimensional, \mathbb{R}^2 se traza una recta que corte a x , se debe contar el número de puntos de la nube a cada lado de la recta e ir rotando hasta obtener el menor número de puntos que quedan a un lado de la misma y ese número entre el número de muestras será la profundidad.

Sigue las propiedades fijadas por Dyckerhoff. Y en el caso del estudio de datos univariantes la profundidad semiespacial es el mínimo entre la función de distribución y la función de supervivencia.



$HD(x, P_x) = \text{minimo} \{F(x), 1 - F(x)\}$, siendo F la función de distribución asociada a P , es decir, $F(x) = P((-\infty, x])$

Un gran inconveniente de usar esta profundidad es la complejidad de cálculo en dimensiones grandes.

4.2. Profundidad simplicial

Esta profundidad es la proporción de símlices que contienen al dato a estudiar dentro de todo aquellos símlices cuyos vértices son $k + 1$ observaciones de la distribución de datos de estudio, ver [9].

Para el caso bivalente \mathbb{R}^2 se debe considerar todos los triángulos cuyos vértices son tres puntos de la muestra y contar cuántos de ellos contienen al punto cuya profundidad es objeto de estudio.

De manera gráfica, se escoge un punto de este semiplano y se construye triángulos cuyos vértices sean observaciones de la muestra dada. Siendo la profundidad simplicial de ese punto escogido la probabilidad de que el punto de estudio se encuentre dentro del polígono construido.

$SD(x, P) = P(x \in S[X_1, X_2, \dots, X_{d+1}])$, $x \in \mathbb{R}^d$, siendo X_1, X_2, \dots, X_{d+1} una muestra de vectores con distribución P y $S[X_1, X_2, \dots, X_{d+1}]$ es el símlice d -dimensional con vértices X_1, X_2, \dots, X_{d+1} .

4.3. Profundidad de Mahalanobis

Es una evolución de la distancia de Mahalanobis [10] entre un punto de estudio y el vector de medias. La transformación de la distancia de Mahalanobis considerada es decreciente y su resultado está acotado entre 0 y 1. Esta función de profundidad tomará su valor máximo, 1, en la media de los puntos a estudiar. Se define como:

$$PMah(x, P) = [1 + (x - \mu)' \Sigma^{-1}(x - \mu)]^{-1}$$

Donde x es el punto cuya profundidad quiere evaluarse, P es la distribución de datos a estudio, μ es la media de la distribución P y Σ^{-1} es la inversa de la matriz de covarianzas.

Suelen usarse definiciones tanto de la media como de la matriz de covarianzas ya que las distribuciones pueden no ser simétricas, estas son:



$$\bar{x} = n^{-1} \sum_{i=1}^n x_i$$

$$S = n^{-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})'$$

El inconveniente más grande en el uso de esta función para el cálculo de la profundidad es la no simetría de algunas funciones de distribución, lo que nos da valores no representativos de estas profundidades estudiadas.

4.4. Profundidad del zonoide

La región central del zonoide de nuestra distribución a estudiar P de nivel α está entre 0 y 1 se define como:

$$ProfZonoide^\alpha(P) = \left\{ \int xg(x)dP(x) : g : \mathbb{R}^d \rightarrow [0,1] \text{ y } \int g(x)dP(x) = 1 \right\}$$

Asumiendo que $\frac{k}{n} < \alpha < \frac{k+1}{n}$, siempre y cuando k esté entre 0 y $(n - 1)$, entonces la definición de la profundidad del zonoide es:

$$ProfZonoide_n^\alpha = co \left\{ \frac{1}{n\alpha} \sum_{i=1}^{i=k} X_{\pi(i)} + \left(1 - \frac{k}{n\alpha} \right) X_{\pi(k+1)} \right\}$$

Siendo co la envolvente convexa y π la permutación desde 1 hasta n . Es decir, la región del zonoide de nivel k/n es un politopo (generalización multidimensional de un polígono) cuyos vértices son promedios de k puntos de la nube.

El punto más profundo de esta distribución será la media y la profundidad de cada punto es el mayor nivel que contiene al punto dentro de su región del zonoide, $D(x; P) = \sup\{\alpha : x \in D^\alpha(P)\}$.

Se muestra un gráfico de la profundidad del zonoide teórico obtenida a través del paquete `dda`, ver [15].

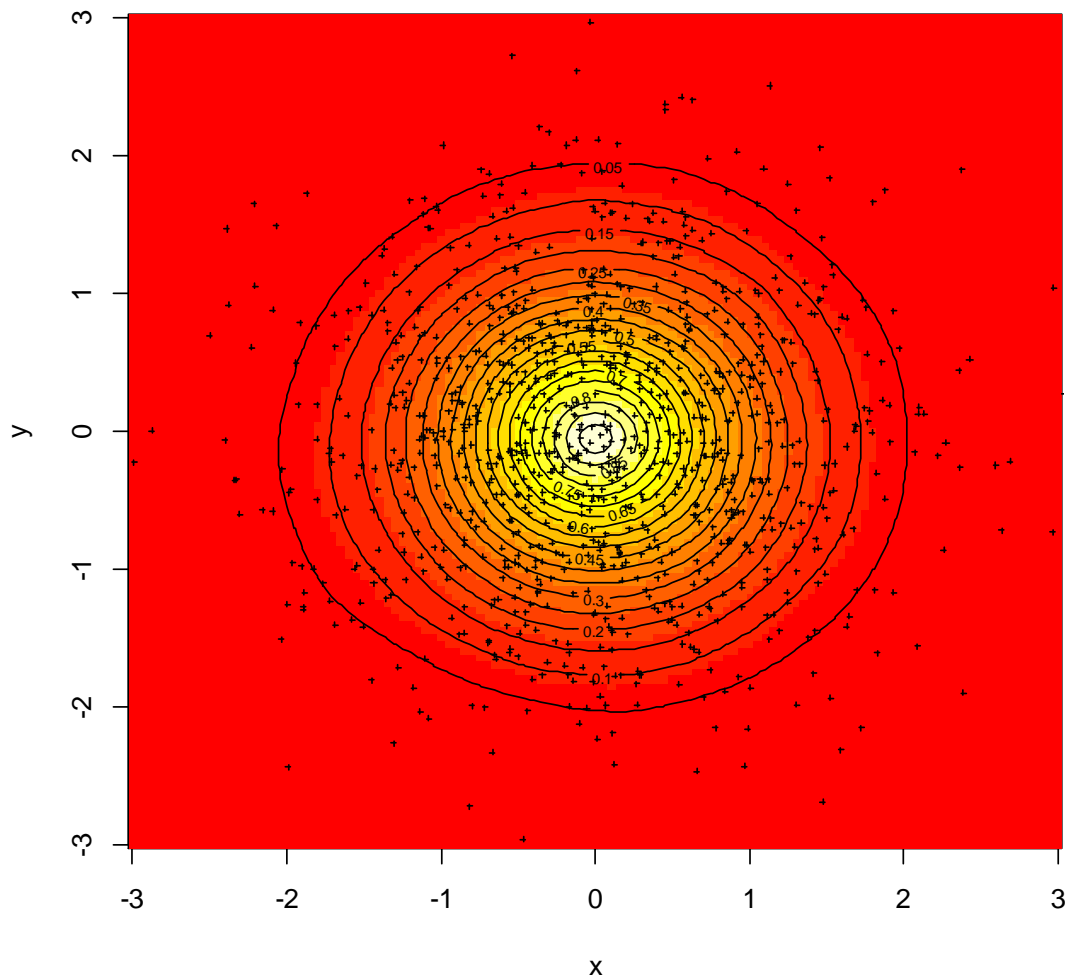


FIGURA 10: “PROFUNDIDAD DEL ZONOIDE TEÓRICA”

El gráfico anterior es una representación de la profundidad del zonoide obtenida a partir 2000 observaciones de datos bidimensionales procedentes de una distribución normal estándar. Los puntos más centrales (color más claro) están en torno al origen de coordenadas y los menos centrales (en color rojo) serán los que tengan una profundidad más próxima a cero, y por tanto más posibilidades de ocasionar inestabilidad en el proceso.

Se construyen los gráficos de control cuyas regiones de control serán las regiones del zonoide de un cierto nivel.



5. Gráficos de control para la media basados en la profundidad del zonoide

Se van a realizar una serie de gráficos de control basados en la profundidad del zonoide, anteriormente explicada.

Estos gráficos sirven para analizar conjuntos de muestras de datos que no tienen por qué ser únicamente unidimensionales y no es necesario tampoco que provengan de una distribución normal, al contrario que en el caso de los gráficos de control citados en el capítulo 3, que únicamente pueden analizar de manera fiable datos unidimensionales y procedentes de distribuciones normales.

Se intentará comprobar si un conjunto de muestras de un proceso industrial se encuentran bajo control; esto es, que los valores de las profundidades obtenidas a partir de estos datos estén por encima de los límites de alarma fijados por nivel de significación α y de los diferentes tamaños muestrales escogidos.

Dependiendo del valor del nivel de significación escogido, se afirma que el proceso es estable aunque alguna de las muestras se encuentre fuera del límite de control, ya que por ejemplo, en el caso de escoger un nivel de significación de 0.05 será aceptable que 1 de cada 20 muestras se encuentre por debajo de los límites de alarma para decir que el proceso es estable.

Para el cálculo de los gráficos de control a partir de las profundidades del zonoide debo recurrir al paquete `ddalpha` del programa estadístico R, ver [15].

El paquete `ddalpha` nos permite calcular una serie de profundidades a través de funciones que vienen predefinidas de antemano, tales como: la profundidad de Mahalanobis, la simplicial, la semiespacial, la profundidad del zonoide, además de otras muchas.

En este caso la función `zonoid.depth(x, data)` proporcionada por el paquete `ddalpha` calculará la profundidad del zonoide de un conjunto de medias muestrales los datos de prueba, con respecto a otro conjunto de datos históricos colocados en una matriz de d columnas, siendo d la dimensión en la que están los datos escogidos.

Las profundidades calculadas de cada muestra son usadas para realizar el gráfico de control, calculando el límite de alarma y la línea mediana y analizar si el proceso a estudio está o no bajo control estadístico.



5.1. Gráfico basado en la profundidad del zonoide para datos unidimensionales

Para la realización de los casos prácticos que se van a exponer se han elegido distribuciones normales de media 35 y desviación típica 0.5, aunque se podría escoger cualquier valor diferente de estos siempre y cuando pertenezcan a una distribución normal, debido a que en este capítulo solo se analiza este tipo de distribución.

El objetivo de la realización de estos gráficos, es observar si las previsiones que se tienen de una serie de muestras, llamadas *pruebas*, se encuentran bajo control estadístico, con respecto a unas muestras de datos, *históricos*, con un límite de alarma fijado, que depende del tamaño muestral y del nivel de significación escogido.

Los datos históricos escogidos y los de previsiones sobre los que se realiza el gráfico, serán unidimensionales y procedentes de una distribución normal.

Se realizan tres gráficos para diferentes tamaños muestrales y niveles de significación (0.0027, 0.01, 0.05), con el fin de evaluar diferentes casos posibles.

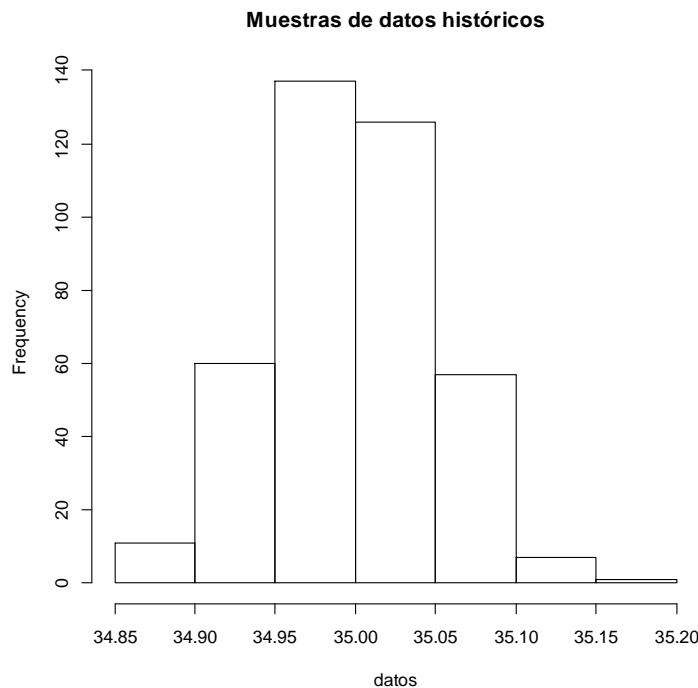


FIGURA 11: "HISTOGRAMA PARA LAS MUESTRAS DE LOS DATOS HISTÓRICOS"

Función en R: `hist (datos, main="Muestras de datos históricos")`



La distribución de *datos* se considerará aceptable siempre y cuando el *p-valor* del test de *Shapiro-Wilk* para esta distribución sea mayor que el nivel de significación escogido.

Se realiza el test de hipótesis a los valores de las muestras de datos históricos.

```
Shapiro-Wilk normality test
data: datos
W = 0.9975, p-value = 0.8075
```

Test de Shapiro-Wilk para los datos históricos (*datos*)

Función en R: `shapiro.test (datos)`

Pertenece a una distribución normal ya que los *p-valor* de los test son mayores que cualquier nivel de significación usado, por lo que se comienza con el análisis.

Se realiza el gráfico de control de la media basado en la profundidad del zonoide.

En esta función se incluye una llamada a otra que calcula la profundidad de las medias muestrales de nuestro conjunto de *pruebas*, dependiente del tamaño muestral escogido, así como una función *plot* que nos dibuja el gráfico de estas medias muestrales recogidas en el vector *profundidades*, calculadas a través de la profundidad del zonoide con respecto al conjunto histórico de datos (*datos*).

En esta función principal será colocado el límite de alarma que no debe ser superado por ninguna muestra para afirmar que el proceso se encuentra bajo control estadístico, son obtenidos a partir de la Tabla 4 que se muestran en el Anexo I. Este límite de alarma será diferente para cada nivel de significación y tamaño muestral del proceso.

Los valores de la profundidad indican la cercanía de la muestra analizada con respecto al máximo, que es 1, cuanto mayor sea este valor, más cercano estará al valor fijado por la especificación, mientras que si este valor es muy próximo a cero lo más probable es que la muestra estará sobrepasando el límite de alarma del gráfico de control, ocasionando una anomalía en el proceso, produciendo inestabilidad en éste.

La línea central que aparece en los gráficos, solo dependerá del tamaño muestral escogido en cada caso, ya que se ha fijado para un valor de $\alpha = 0.5$ con el fin de que aproximadamente la mitad de las muestras tomadas durante el proceso se encuentren por encima y la otra mitad por debajo de la línea, para que el gráfico sea más visual, aunque se podría haber fijado otro valor de α .



La función de profundidad que calcula las profundidades de los datos previstos con respecto a los históricos es $zdepth(s, x)$, siendo x el conjunto de las medias muestrales de los datos de prueba, y s será el conjunto de datos históricos.

Para el cálculo de la media del conjunto de datos que existen en *pruebas* que el conjunto es una matriz con un número de columnas igual al tamaño muestral escogido que se almacena en una variable llamada *mediasx* y será introducida en la función `vzdepth (mediasx, datoshistoricos)`, para obtener las profundidades de estas medias con respecto a los datos *históricos*.

Se muestran tres gráficos con diferentes tamaños muestrales y diferentes niveles de significación para comprobar si el proceso es estable en diferentes escenarios.

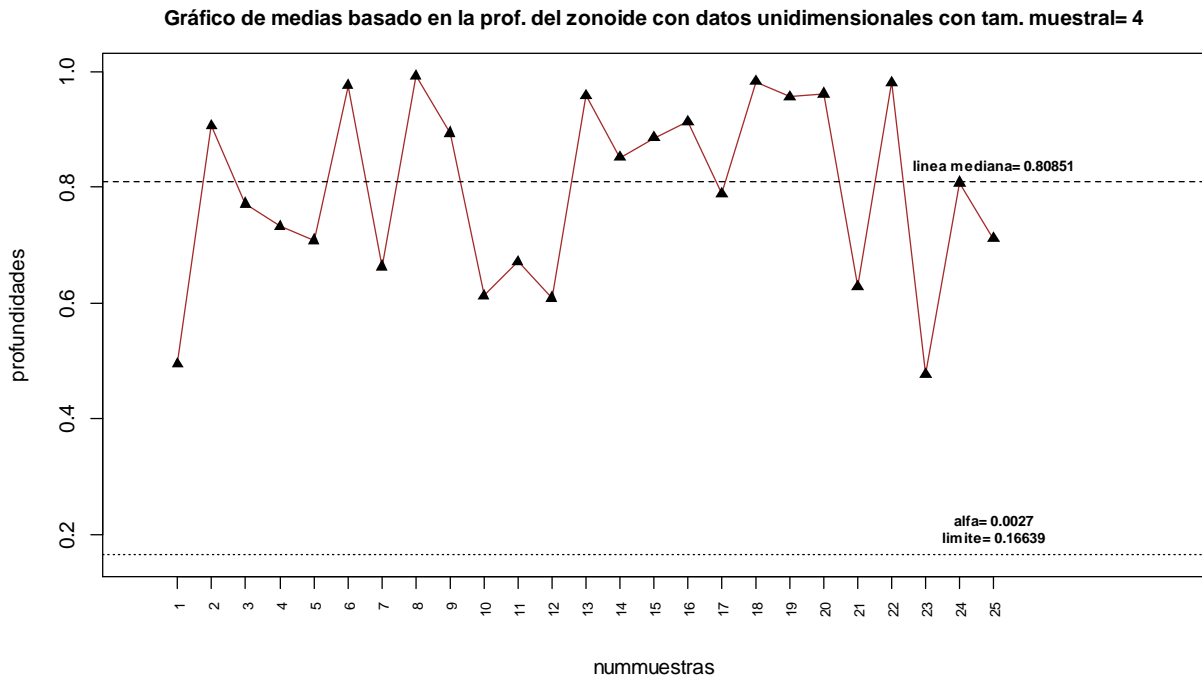


FIGURA 12: “GRÁFICO DE CONTROL BASADO EN LA PROFUNDIDAD DEL ZONOIDE ($\alpha=0.0027$ Y $N=4$)”

Función en R: `graficounidimensional (pruebas, datos, 4, 0.0027)`



En el gráfico aparecen 25 muestras tomadas de los datos de prueba o previsiones, así como sus profundidades con respecto a los datos históricos.

Se observa que como se ha dicho antes la mitad de las muestras tomadas se encuentran por encima de la línea mediana del gráfico y la otra mitad se encuentran por debajo (11 frente a 9). Además se afirma que ninguna de las muestras se encuentra por debajo del límite de alarma fijado, en este caso 0.16639, por lo que diremos que el proceso se encuentra bajo control para este nivel de significación y tamaño muestral.

El límite de alarma anteriormente citado dependerá del nivel de significación usado en el análisis, así como del tamaño muestral elegido. En los demás gráficos el límite de alarma y la línea central no tienen los mismos valores que en este caso. Dichos valores están expuestas en la Tabla 5 del Anexo I.

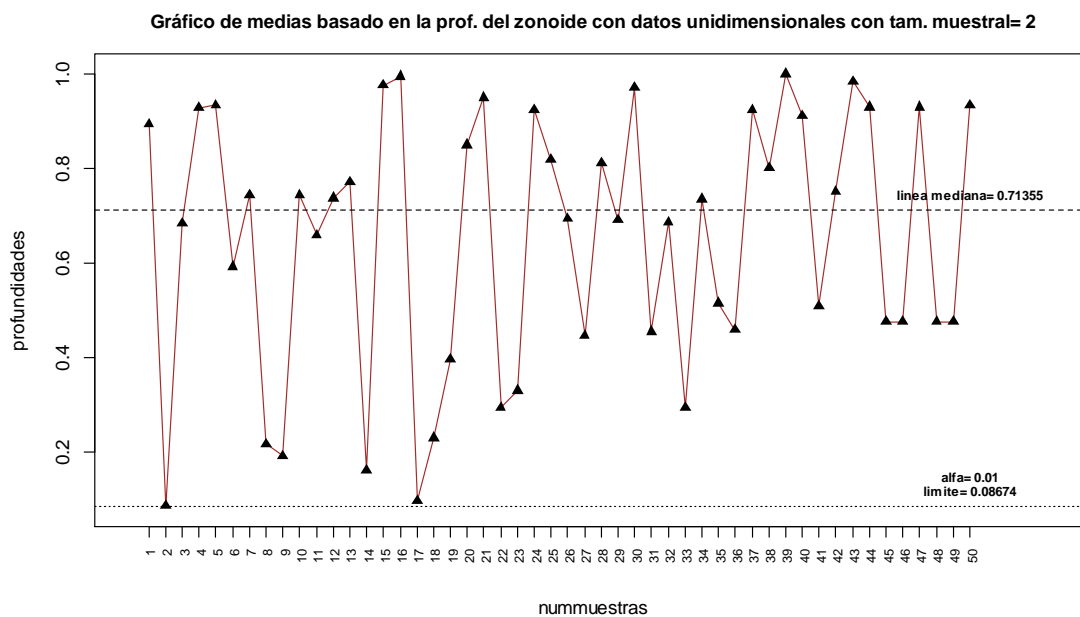


FIGURA 13: “GRÁFICO DE CONTROL BASADO EN LA PROFUNDIDAD DEL ZONOIDE ($\alpha=0.01$ Y $N=2$)”

Función en R: `graficounidimensional (pruebas, datos, 2, 0.01)`

Al igual que ocurría anteriormente, este gráfico de control también nos indica que el proceso estudiado sigue siendo estable, aunque hayamos cogidos valores del tamaño muestral y del nivel de significación diferentes.

Es estable ya que ningún valor de las muestras queda por debajo del límite de alarma considerado como aceptable.

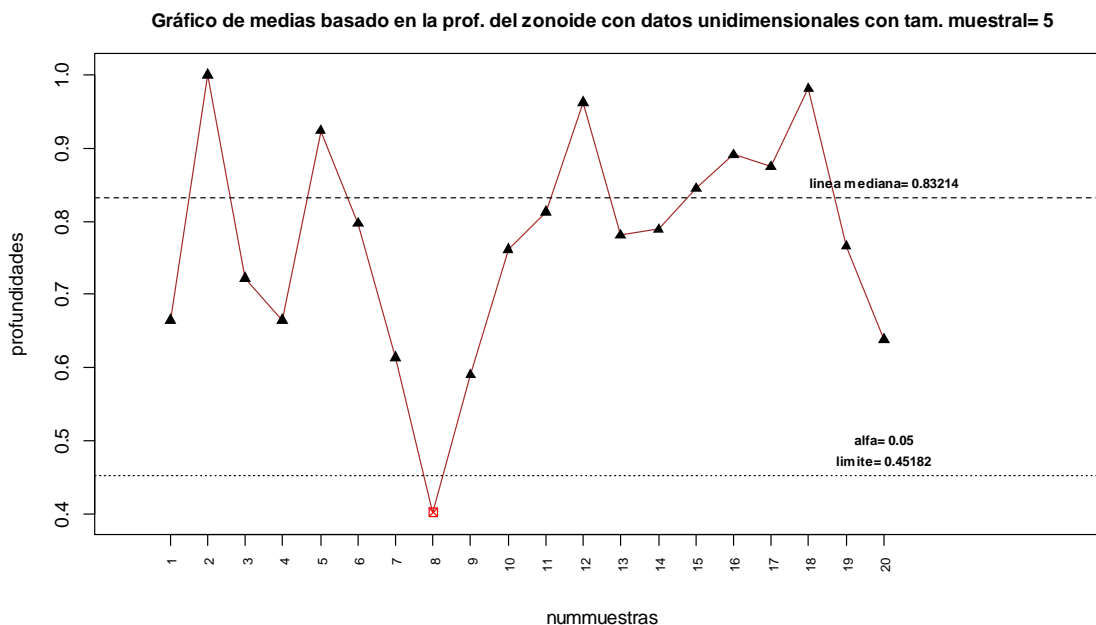


FIGURA 14: “GRÁFICO DE CONTROL BASADO EN LA PROFUNDIDAD DEL ZONOIDE ($\alpha = 0.05$ Y $N=5$)”

Función en R: `graficounidimensional (pruebas, datos, 5, 0.05)`

En este caso, la muestra número 8 se encuentra por debajo del límite de control de nuestro gráfico; sin embargo, el proceso sigue estable y bajo control, ya que es esperable que con un nivel de significación de 0.05, 1 de cada 20 muestras se encuentre fuera de este límite.

A la vista de los resultados se afirma que el proceso se encuentra bajo control estadístico para todos los niveles de significación y tamaños muestrales.

5.2. Estudio del gráfico de control basado en la profundidad del zonoide para datos en dos dimensiones

Los datos escogidos históricos y los de prueba, sobre los que se realiza la profundidad del zonoide, serán bidimensionales, y habrá tres gráficos para diferentes tamaños muestrales (desde 2 hasta 10) y niveles de significación (0.0027, 0.01, 0.05).

Al igual que en el caso anterior, es necesario el paquete *ddalpha* del programa R, con el fin de poder utilizar la función predefinida `zonoid.depth(x, data)` que permite obtener la profundidad de las muestras de datos de prueba frente a la serie de datos históricos.



Tanto los datos históricos como los datos de prueba proceden de distribuciones normales y son obtenidos a través de la función `rnorm(400, 35, 0.5)` para los históricos y se almacenan en una matriz de dos columnas, ya que el número de columnas será el número de dimensiones de los datos, y lo mismo ocurrirá con los datos de prueba: `rnorm(100, 35, 0.5)`.

Se dispone de un total de 200 datos bidimensionales históricos y de 50 datos bidimensionales de prueba. El número de muestras variará en función del tamaño muestral escogido.

Se presentan los datos históricos en la siguiente figura.

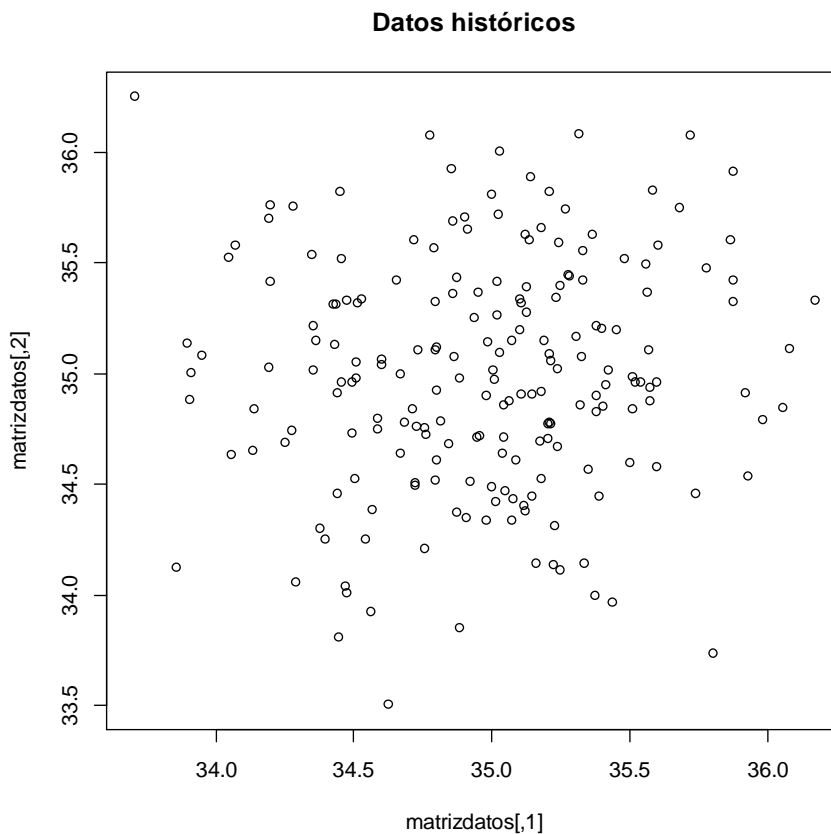


FIGURA 15: "DATOS HISTÓRICOS"

Para la realización del análisis se debe calcular las medias muestrales de las muestras seleccionadas de la matriz de pruebas a través de la función `mediasmuestrales(matrizpruebas, k)` que se almacena en el vector *medias*, con el fin de compararlo con los datos históricos.

Para obtener las profundidades de las medias muestrales es usada la función de la profundidad del zonoide, `zonoid.depth(medias, matrizd)` almacenando sus resultados en la variable *profundidades*, y una vez hecho esto, se grafica este



vector de profundidades. El número de muestras elegido se coloca en el eje de abscisas, y en el eje de ordenadas se indica el valor de la profundidad de cada muestra, desde 0 hasta 1.

Se ejecuta la función `graficobivariantes (matrizp, matrizd, alfa, k)` que imprimirá por pantalla el gráfico de las profundidades de las medias muestrales de los datos de *prueba* con un límite de alarma que dependerá del nivel de significación y del tamaño muestral escogido.

Se presentan tres gráficos usando las mismas distribuciones de datos, pero modificando el nivel de significación y el tamaño muestral en cada uno de ellos, con objeto de detectar si nuestro proceso teórico es estable para cualquier valor del nivel de significación o del tamaño muestral escogido.

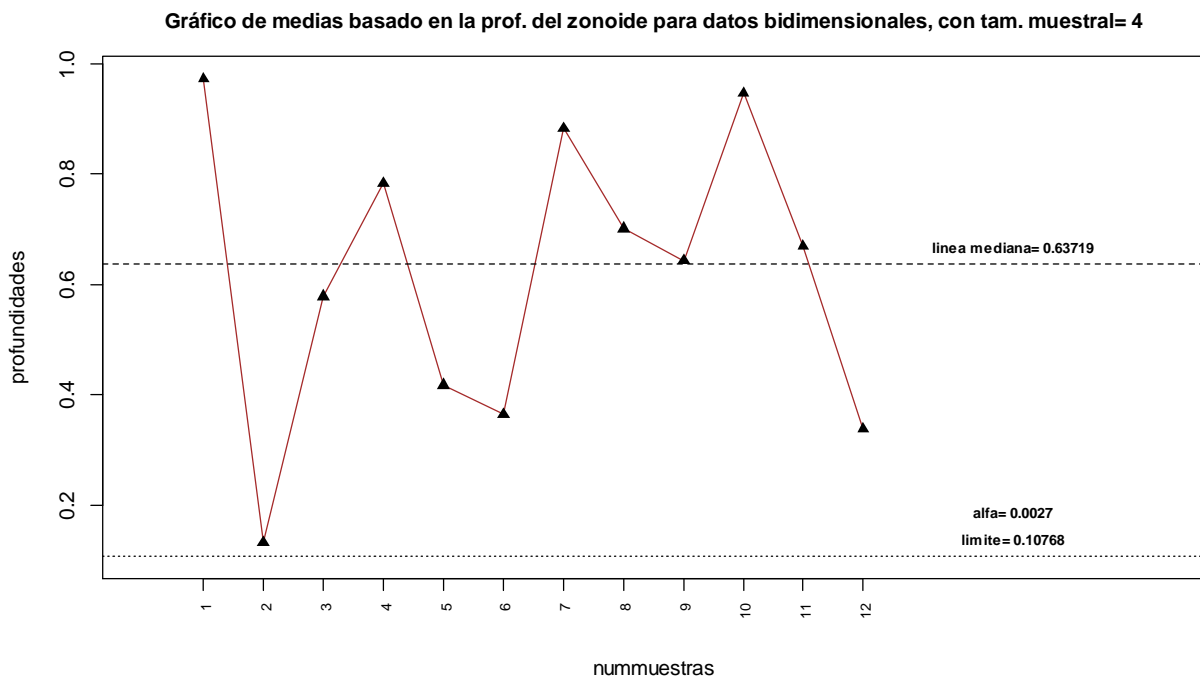


FIGURA 16: “GRÁFICO DE CONTROL BASADO EN LA PROFUNDIDAD DEL ZONOIDE, DATOS BIDIMENSIONALES ($\alpha = 0.0027, N=4$)”

Función en R: `graficobivariantes (matrizpruebas, matrizdatos, 4, 0.0027)`

Se observa en el gráfico anterior que ninguna muestra queda por debajo del límite, que es 0.10768, por lo que el proceso se encuentra estable para este nivel de significación y tamaño muestral.



Gráfico de medias basado en la prof. del zonoide para datos bidimensionales, con tam. muestral= 2

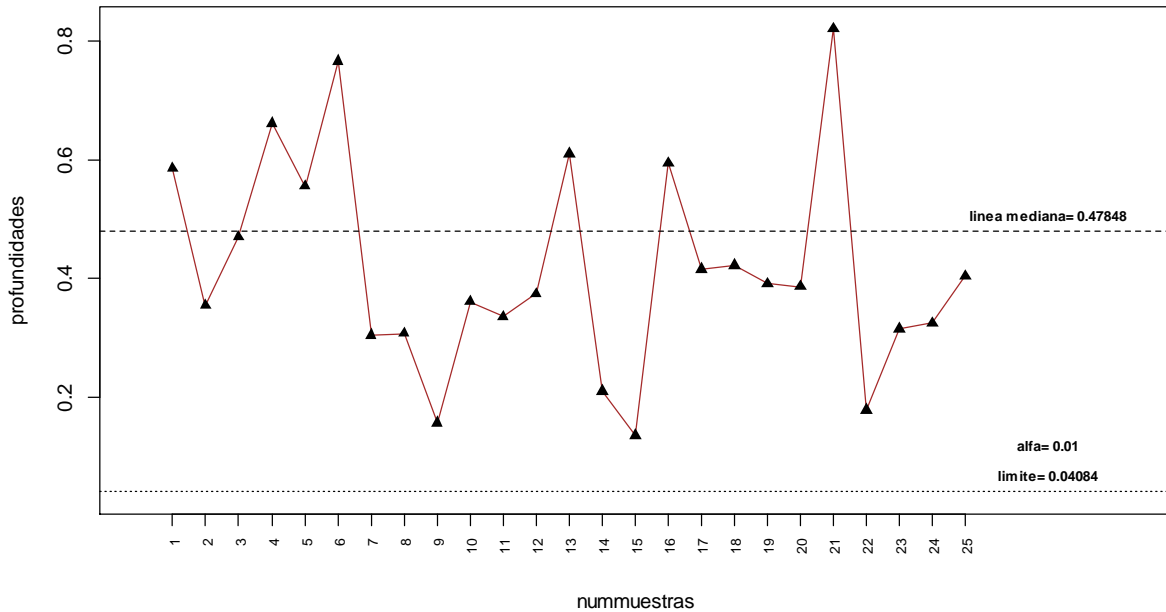


FIGURA 17: “GRÁFICO DE CONTROL BASADO EN LA PROFUNDIDAD DEL ZONOIDE CON DATOS BIDIMENSIONALES ($\alpha=0.01, N=2$)”

Función en R: `graficobivariantes(matrizpruebas, matrizdatos, 2, 0.01)`

También se observa que todas las muestras se encuentran por encima del límite de alarma del gráfico, 0.04084, por lo que afirmaremos que el proceso se encuentra bajo control y es estable para estas especificaciones.

Gráfico de medias basado en la prof. del zonoide para datos bidimensionales, con tam. muestral= 2

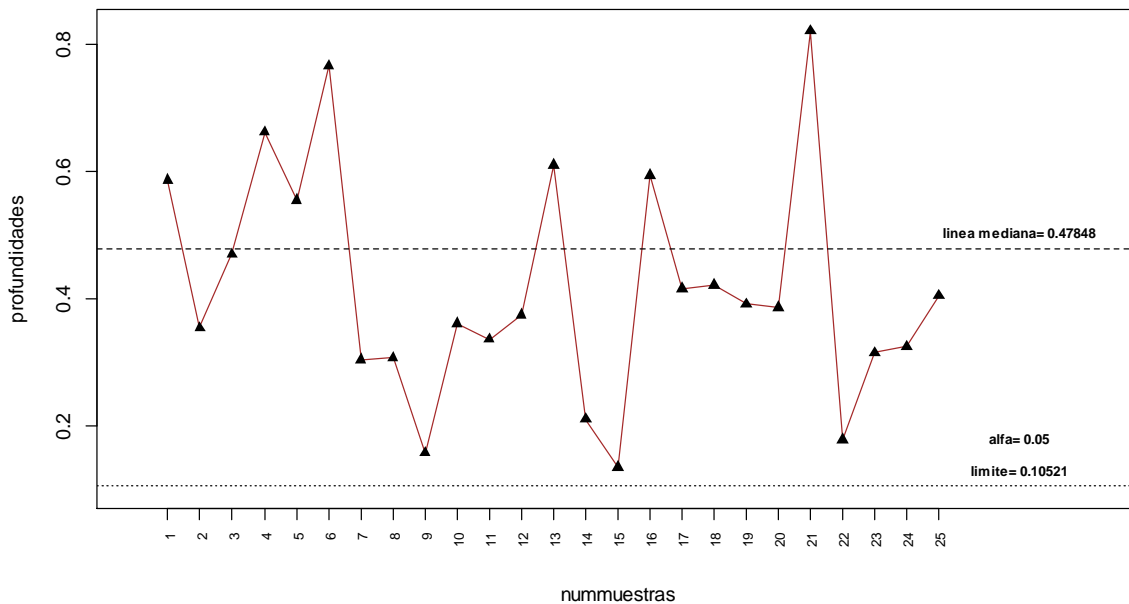


FIGURA 18: “GRÁFICO DE CONTROL BASADO EN LA PROFUNDIDAD DEL ZONOIDE CON DATOS BIDIMENSIONALES ($\alpha=0.05, N=2$)”

Función en R: `graficobivariantes(matrizpruebas, matrizdatos, 2, 0.05)`



El proceso, al igual que anteriormente, sigue estando bajo control estadístico ya que todas las muestras se encuentran de los límites de alarma, por lo que afirmamos que el proceso es estable y no aparecen anomalías a la vista.

Se afirma que el proceso es estable para todos los niveles de significación y tamaños muestrales llevados a estudio.

5.3. Estudio del gráfico de control basado en la profundidad del zonoide para datos en tres dimensiones

Para el último apartado, los datos históricos son escogidos de la misma forma que anteriormente, es decir, son datos procedentes de distribuciones normales.

Para los datos históricos se utiliza la función `rnorm(810, 35, 0.5)` que proporciona el programa R y los datos son almacenados una matriz de 3 columnas ya que se trata de datos en tridimensionales, y lo mismo ocurrirá con la matriz de datos de prueba, cuyos datos son extraídos a través de la función `rnorm(210, 35, 0.5)`, debido a que también pertenecen a una distribución normal.

Los datos no pueden ser representados a través de un histograma ni tampoco a partir de un diagrama, ya que al estar en tres dimensiones, no hay ningún medio para su representación gráfica, a diferencia de los dos casos anteriores, donde estos se representan porque son datos unidimensionales o bidimensionales.

Se dispone de un total de 270 datos históricos que compararemos con 70 datos previstos.

Realizaré tres gráficos con diferentes tamaños muestrales y niveles de significación con la misma función del capítulo anterior, añadiendo una línea de código debido a que se ha añadido una dimensión más en los datos.

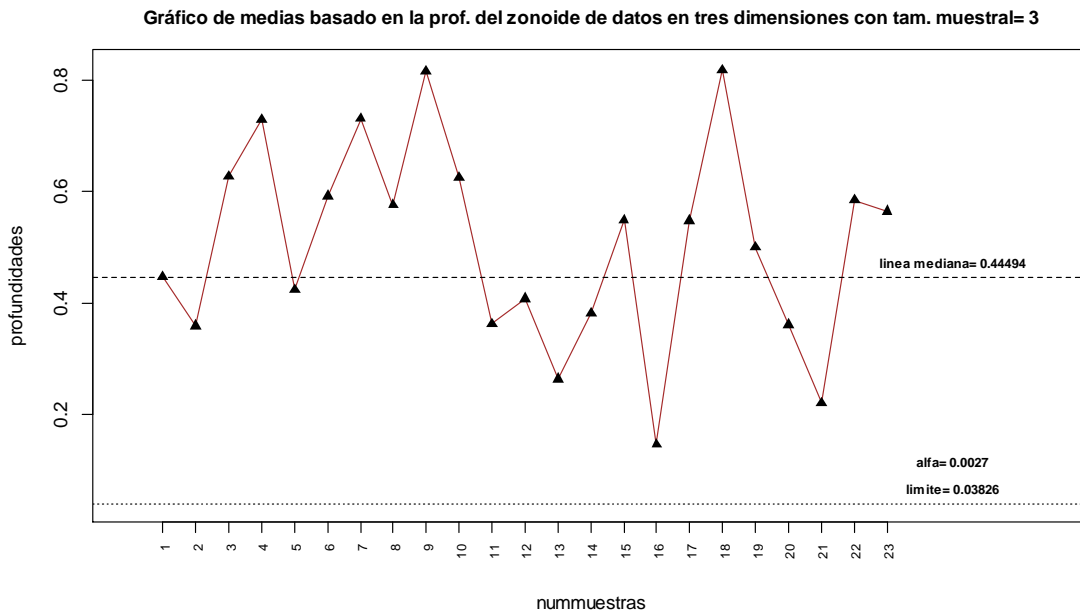


FIGURA 19: “GRÁFICO DE CONTROL BASADO EN LA PROFUNDIDAD DEL ZONOIDE CON DATOS TRIDIMENSIONALES ($\alpha=0.0027, N=3$)”

Función en R: `graficotrivariantes(matrizpruebas, matrizdatos, 3, 0.0027)`

Podemos ver que todas las muestras se encuentran por encima del límite de alarma impuesto, por lo que el proceso es estable para este nivel de significación y tamaño muestral usado.

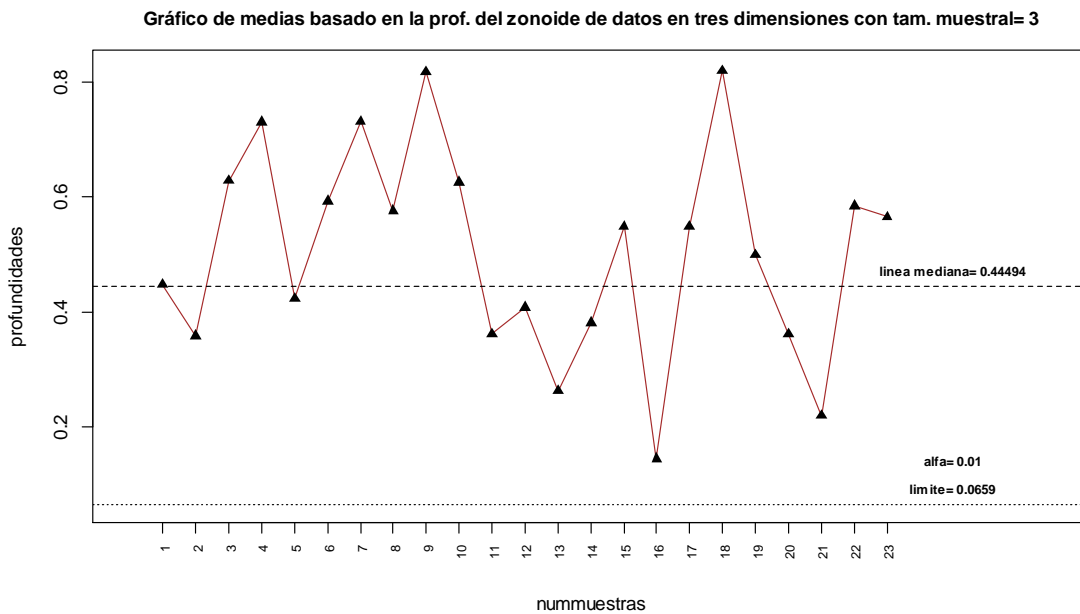


FIGURA 20: “GRÁFICO DE CONTROL BASADO EN LA PROFUNDIDAD DEL ZONOIDE CON DATOS TRIDIMENSIONALES ($\alpha=0.01, N=3$)”

Función en R: `graficotrivariantes(matrizpruebas, matrizdatos, 3, 0.01)`



Todos los valores de las profundidades de las muestras se encuentran por encima del límite por lo que afirmamos que el proceso se encuentra bajo control para este nivel de significación y tamaño muestral, al igual que hemos visto anteriormente.

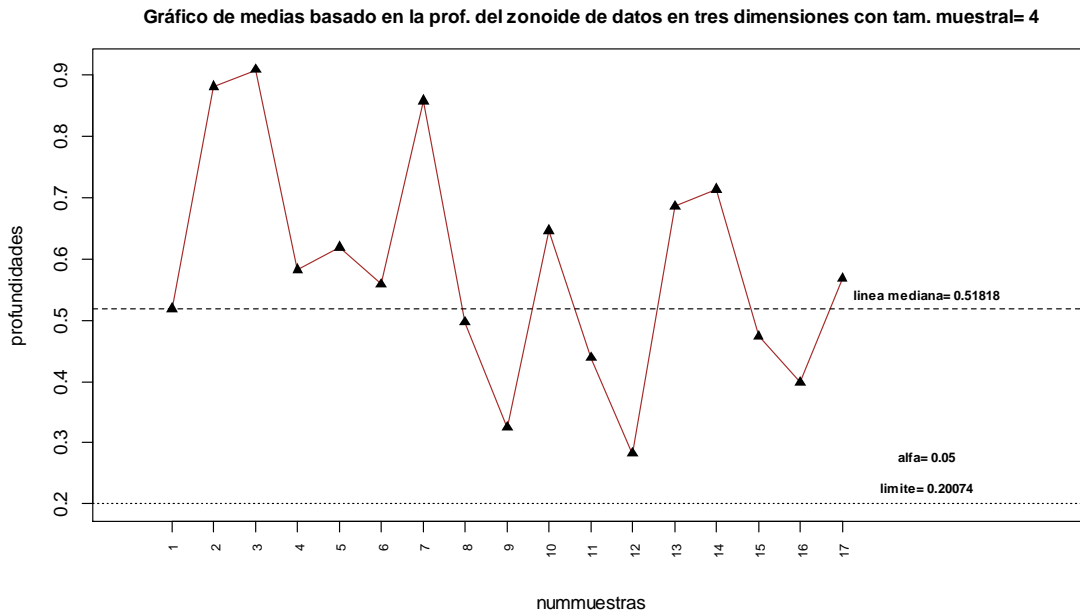


FIGURA 21: “GRÁFICO DE CONTROL BASADO EN LA PROFUNDIDAD DEL ZONOIDE CON DATOS TRIDIMENSIONALES ($\alpha=0.05, N=4$)”

Función en R: `graficotrivariantes(matrizpruebas, matrizdatos, 4, 0.05)`

El límite de alarma y la línea mediana se han obtenido a través de los datos que se encuentran en la tabla 6 del Anexo I.

Al igual que ocurre en los dos gráficos anteriores, todas las muestras se encuentran por encima del límite de alarma, por lo que se afirma que después del análisis para diferentes valores de α y tamaños muestrales, el proceso se encuentra bajo control para todos los niveles y tamaños muestrales.



6. Profundidad de localización y escala

Se quiere analizar una serie de datos a través del gráfico de control basado en la profundidad de localización y escala, y observar si también es posible llegar a conclusiones lógicas, pero para ello primero se debe definir el concepto de profundidad de localización y escala.

La profundidad de localización y escala es diferente a todas las que hemos explicado anteriormente, ya que las otras identificaban el grado de centralidad de un punto con respecto a una distribución dada. Sin embargo esta profundidad sirve para conocer cuándo un punto es idóneo para representar un parámetro de localización de nuestra distribución de estudio, es decir, puede ser un *nonfit*.

El elemento *nonfit* es aquel punto que sirve para concretar un parámetro de la distribución de estudio. Ivan Mizera [20] explica que la profundidad de localización es la menor proporción de observaciones que hay que eliminar para que un parámetro de estudio sea un *nonfit*. Un punto será un *nonfit* cuando se encuentre en el exterior de la envolvente convexa esperada de la distribución.

Se puede calcular la profundidad de localización y escala gracias a los teoremas que expusieron I. Mizera y C.H. Müller en los que extienden la profundidad a parámetros bidimensionales y se calcula a partir de modelos de distribución unidimensionales, ya que se basan en estos modelos de una dimensión, ver [20].

Profundidad de localización y escala: $D_{LS}((m, s), P)$,

Siendo la profundidad de localización y escala de la muestra escogida, la mayor fracción de puntos de la nube tal que m es su promedio y s es su desviación típica, se puede aproximar a la función del zonoide como:

$$D_{LS}((m, s); P_x) = D((m, s^2 + m^2); P_{x,x^2}) \quad [23]$$

Siendo $s^2 + m^2$ el promedio de los cuadrados, D se refiere a la profundidad del zonoide anteriormente citada.

Por lo que sí es posible calcular la profundidad del zonoide bidimensional, podremos calcular la profundidad de localización y escala de cualquier tipo de distribución.

Al igual que el gráfico basado en la profundidad del zonoide, en este también aparece un límite de alarma y una línea mediana. Estos dependerán de los valores del tamaño muestral y del nivel α escogido en el análisis.

Los valores de éstos, se encuentran en la Tabla 7 del Anexo I, ya que como hemos dicho, la profundidad de localización y escala se puede aproximar al cálculo de la profundidad del zonoide.



Ejemplo práctico

Para realizar el ejemplo práctico se escogen los mismos datos que en el gráfico de control de la media basado en la profundidad del zonoide para datos bidimensionales.

Son usadas dos matrices, una para los históricos y otra para los datos de prueba.

En la primera columna de las matrices se colocan los datos a estudiar, mientras que en la segunda, son colocados estos mismos datos elevados al cuadrado, ya que es lo que indica la función anteriormente mencionada.

Se calcula las medias muestrales de cada una de las muestras escogidas en las *pruebas* a través de la función `mediasmuestrales(matrizpruebas,k)` introduciendo los datos previstos (*pruebas*) y el tamaño muestral elegido.

Se usa la función creada en el programa estadístico R:

`graficolocalizacionescala(matrizpruebas,matrizdatos,k,alfa)`, para realizar el gráfico de control.

Dentro de la función anteriormente citada se calcula el límite de alarma del gráfico que dependerá del tamaño muestral y del nivel de significación *alfa*. Además del valor de la línea mediana del gráfico, que solo dependerá del tamaño muestral escogido, porque igual que anteriormente, el nivel de significación para el cálculo de esta 0.5.

Se calcula la profundidad de localización y escala a partir de la misma función usada en el gráfico de la profundidad del zonoide para datos bidimensionales: `depth.zonoid(medias, matrizdatos)`, y se acumulan estos valores en la variable *profundidades*.

La función termina realizando un `plot` en el que se coloca el número total de muestras a estudiar, en el eje X, y el valor de cada una de las profundidades asociadas a cada muestra en el eje Y, incluyendo también el valor límite a través del que analizamos si nuestro proceso se encuentra bajo control (un porcentaje específico de valores se encuentra por encima de éste).

Una vez explicado el proceso de manera teórica, se muestran varios gráficos con el objetivo de analizar si el proceso se encuentra o no bajo control.

Cada uno de los puntos que aparecen en los gráficos se obtiene comparando la media y la desviación típica de cada una de las muestras previstas frente a los datos históricos.

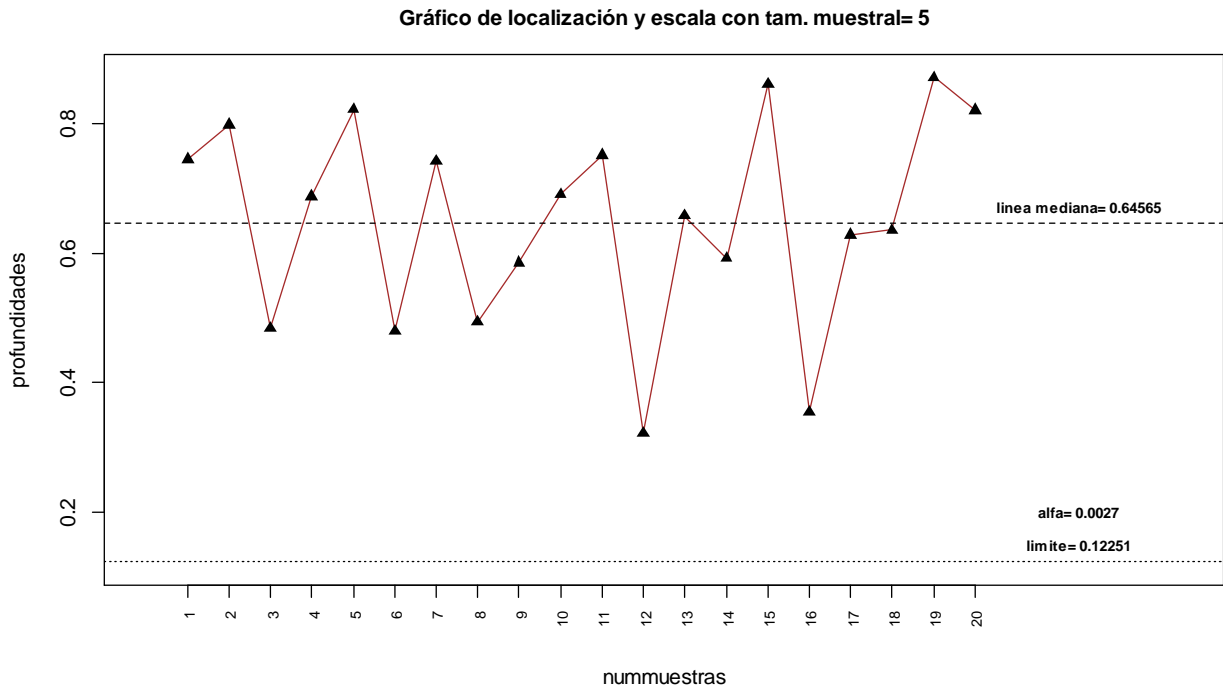


FIGURA 22: “GRÁFICO DE PROFUNDIDAD DE LOCALIZACIÓN Y ESCALA (ALFA = 0.0027, K = 5)”

Función en R: `graficolocalizacionescala(matrizpruebas, matrizdatos, 5, 0.0027)`

Como se ha dicho antes, en el gráfico se puede apreciar que la mitad de las muestras usadas para el análisis se encuentran por encima de la línea mediana, y la otra mitad por debajo, simplemente para que el gráfico sea más visual.

El valor del límite de alarma se encuentra fijado para este tamaño muestral y nivel de significación en la tabla 7 del anexo I ya que en el proceso se analizan datos en dos dimensiones, en este caso el valor de la línea mediana es 0.64565 y el del límite de alarma es 0.12251, correspondientes en la tabla a un valor de tamaño muestral 5 y nivel $\alpha = 0.0027$.

El valor de las profundidades de todas las muestras obtenidas se encuentran por encima del límite de alarma, por lo que este proceso será estable para este nivel de significación y tamaño muestral.

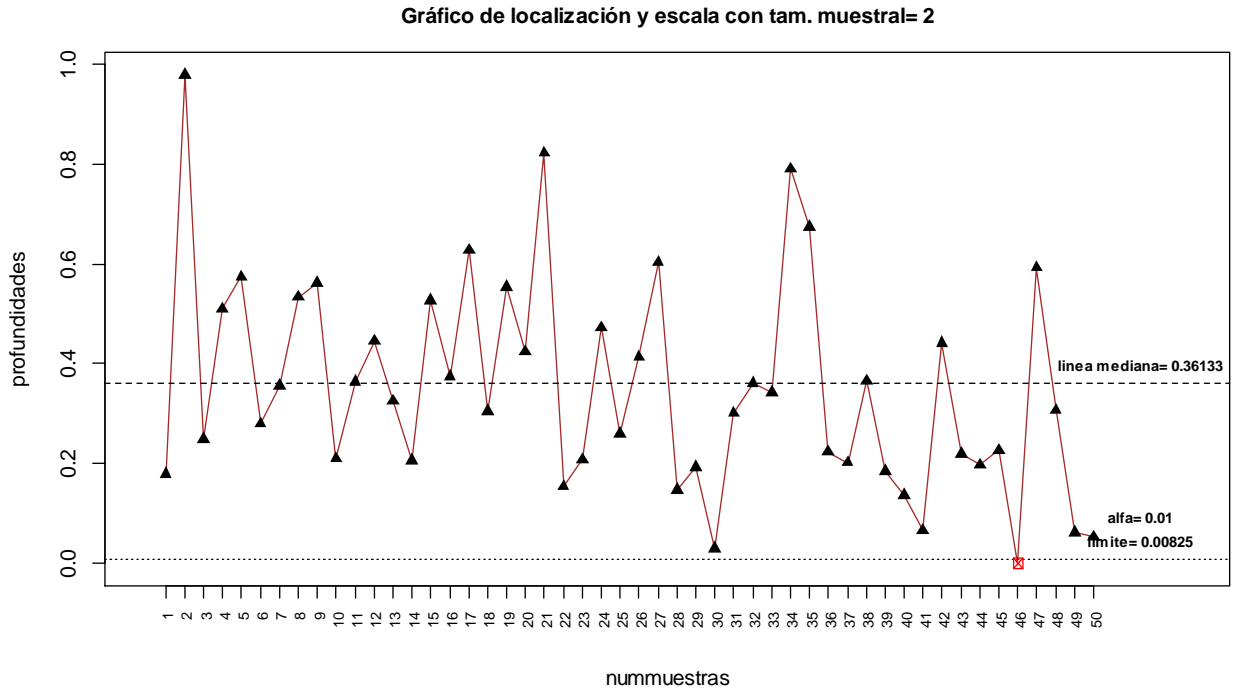


FIGURA 23: “GRÁFICO DE PROFUNDIDAD DE LOCALIZACIÓN Y ESCALA (ALFA = 0.01, K = 2)”

Función en R: `graficolocalizacionescala(matrizpruebas, matrizdatos, 2, 0.01)`

En el gráfico de arriba se observa que una de las muestras sobrepasa el límite de alarma, la muestra número 46, pero esto no quiere decir que el proceso sea inestable, ya que es esperable porque el nivel de significación es más alto que en el caso anterior (0.01 frente a 0.0027), y esperamos que una o dos muestras de cada 100 estén fuera de control, y en este caso hay una fuera de control de 50 muestras, por lo que a tenor de lo expuesto, se puede afirmar que el proceso se encuentra bajo control estadístico para estos niveles de análisis escogidos.

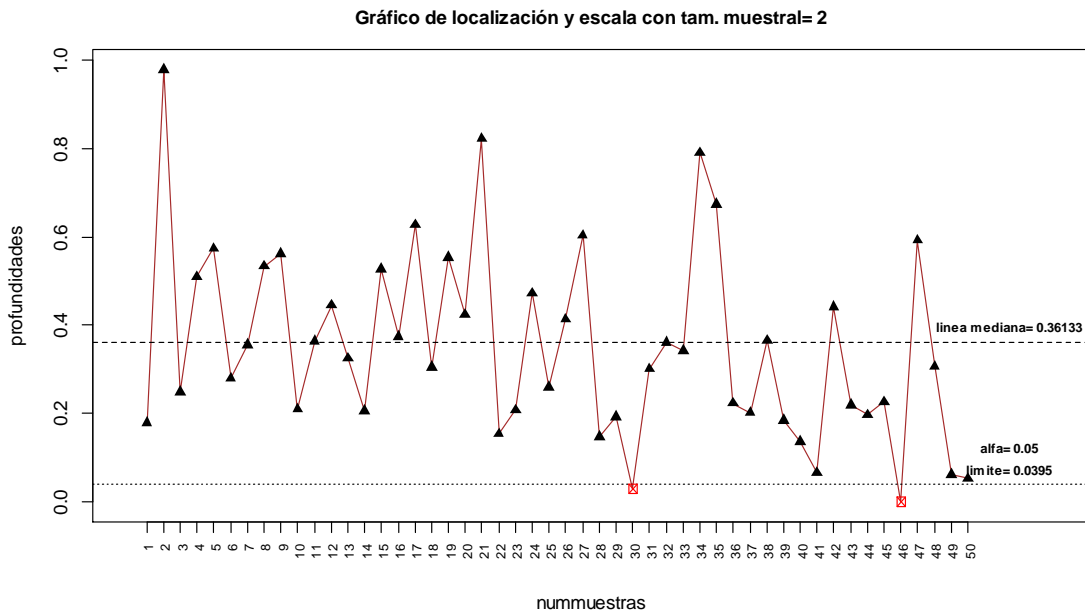


FIGURA 24: “PROFUNDIDAD DE LOCALIZACIÓN Y ESCALA (ALFA = 0.05, K = 2)”

Función en R: `graficolocalizacionescala(matrizpruebas, matrizdatos, 2, 0.05)`

Al igual que en el caso anterior, en el gráfico aparecen dos muestras por debajo del límite de alarma, la 30, la y la 46, pero el proceso se encuentra estable ya que a partir de ese nivel de significación escogido es sabido que 5 o 6 de cada 100 muestras estarán por debajo de ese límite y en este caso aparecen 2 de 50 por lo que consideramos que el proceso es estable aunque estas dos muestras se encuentren por debajo de los límites de control, ya que estos errores son admisibles para este nivel de significación escogido.

Por tanto el proceso que se ha estudiado será estable para todos y cada uno de los diferentes niveles de significación usados durante el análisis.



7. Comparativa para datos unidimensionales

En un proceso industrial es necesario conocer qué tipo de gráfico de control es más útil en cuanto a analizar los datos de dicho proceso, por eso en éste y en el siguiente capítulo se realiza una comparativa entre los tres tipos de gráficos estudiados y se ejecuta un análisis de cuál es más óptimo para el caso de que los datos tengas diferentes dimensiones.

Se realiza esta comparativa, ya que estos gráficos anteriores han sido ejecutados, imponiendo que los datos del proceso pertenecen a una distribución normal, pero esto no siempre se cumple.

Por tanto, se muestra si todos los gráficos de control son útiles para el caso de analizar datos unidimensionales que pertenezcan a una distribución normal o no.

Se usan los mismos datos para los tres gráficos de control obtenidos a través de una distribución exponencial de 83 muestras de tamaño 3 con un total de 249 datos y serán analizados para un nivel de significación $\alpha = 0.05$, en todos los casos, con el fin de evaluar de manera equitativa los tres casos.

El primer gráfico presenta el gráfico de control de las medias de Shewhart, el segundo es el gráfico de control de la media basado en la profundidad del zonoide, y el tercero es el gráfico de medias basado en la profundidad de localización y escala.

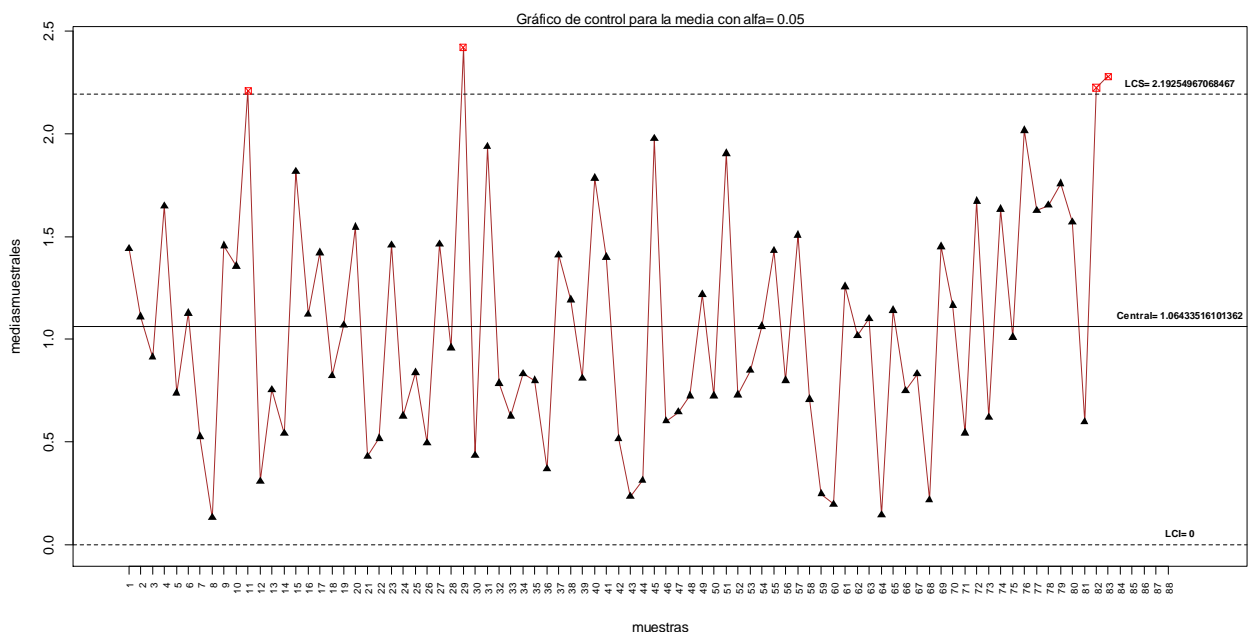


FIGURA 24: “GRÁFICO DE CONTROL DE SHEWHART BASADO EN LA MEDIA PARA DATOS EXPONENCIALES”

Función en R: `graficomedias(datos1, 3, 0.05)`

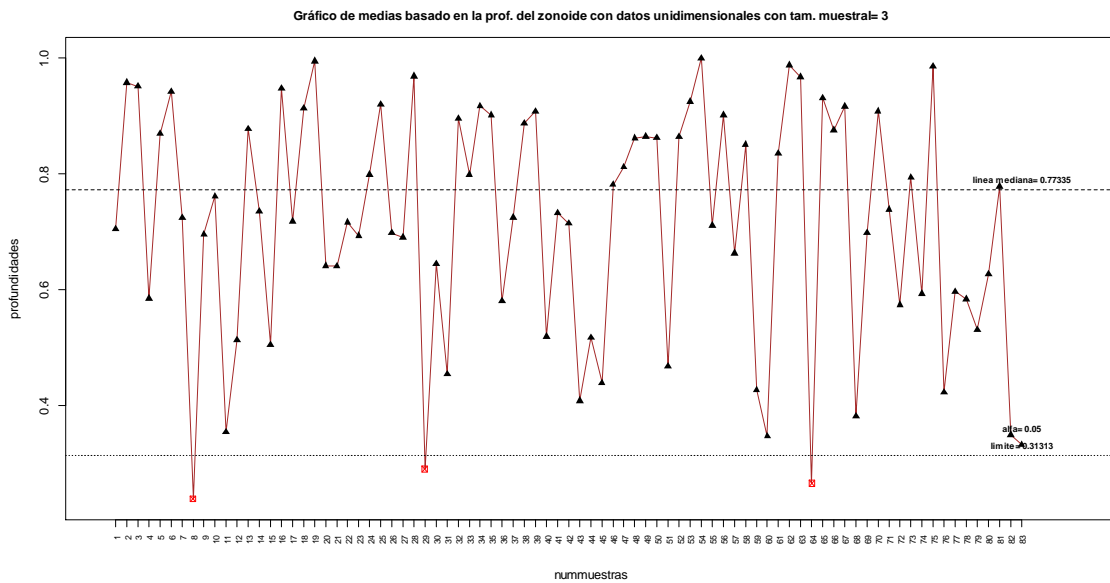


FIGURA 25: "GRÁFICO DE CONTROL DE LA MEDIA PARA DATOS EXPONENCIALES, PROFUNDIDAD DEL ZONOIDE"

Función en R: `graficounidimensional(datos, 3, 0.05)`

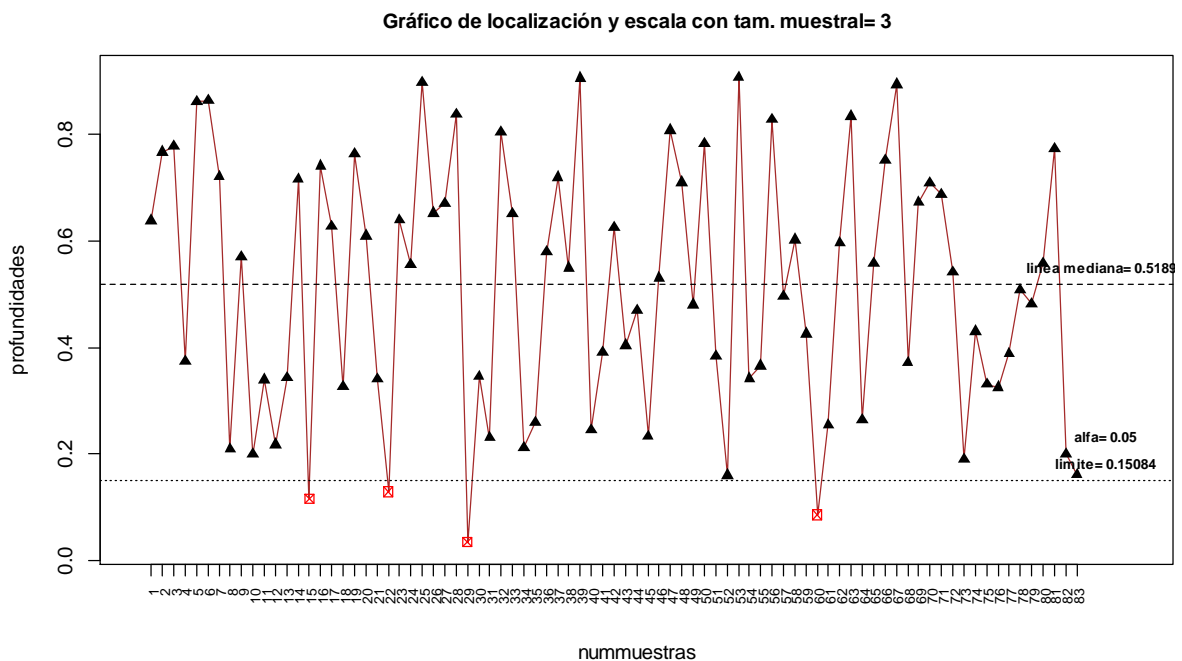


FIGURA 26: "GRÁFICO DE LA MEDIA BASADO EN LA PROFUNDIDAD DE LOCALIZACIÓN Y ESCALA, DATOS EXPONENCIALES"

Función en R: `graficolocalizacionescala(matrizdatos, matrizdatos, 3, 0.05)`



Como se observa en el primer gráfico aparecen 4 muestras en el exterior de los límites de control, por lo que podríamos afirmar que el proceso se encuentra fuera de control, pero no para este nivel de significación escogido ya que se espera que 5 de cada 100 muestras se encuentren fuera de control, por lo que seguimos considerando estable el proceso.

En el segundo gráfico, analizando los mismos datos, se encuentran 3 muestras fuera de control, por lo que se afirma el conjunto de las muestras se encuentra estable, ya que al igual que anteriormente el nivel de significación escogido es el mismo.

Y por último en el tercer gráfico de control vuelven a aparecer parecen cuatro muestras fuera de control, pero no coinciden con las que aparecían en los dos gráficos anteriores, por lo que aunque devuelva una estabilidad manifiesta en el proceso, esto es erróneo ya que considera válidas muestras que se encuentran fuera de control en los gráficos anteriormente citados.

Después de analizar esto, se llega a la conclusión de que cuando la distribución de los datos es no normal, los métodos tradicionales, como el gráfico de control de Shewhart a la hora de analizar datos unidimensionales puede detectar anomalías, cuando estas si existen, al igual que el gráfico basado en la profundidad del zonoide, mientras que el gráfico basado en la profundidad de localización y escala no detecta las anomalías reales pero si detecta como anomalías muestras perfectamente válidas.

También se observa que la región que el zonoide fabrica para los datos se adapta mejor a los datos del proceso que cualquiera de las otras regiones fabricadas por los otros métodos analizados, tanto profundidad de localización y escala como el método de Shewhart.

Por tanto, para el caso de datos unidimensionales procedentes de distribuciones no normales los métodos de control más fiables para su análisis son los gráficos basados en la profundidad del zonoide, ya que aunque haya mucha variación en la media de las diferentes muestras, este método es capaz de ajustarse más a los datos, presentando una mayor eficiencia y fiabilidad que cualquiera de los otros dos.



8. Comparativa entre profundidad zonoide y T^2 de Hotelling para datos bidimensionales

En este capítulo se analiza qué ocurre cuando las distribuciones de datos bidimensionales proporcionadas para el estudio no son normales, sino que pertenecen a otro tipo de distribución, ya sea Cauchy, exponencial, binomial o cualquier otra.

Se compara el gráfico de la media basado en la profundidad del zonoide para datos bivariantes, con el obtenido a través del gráfico de control T^2 de Hotelling.

Pero antes de realizar esto, se necesita saber qué es el gráfico de control basado en el método de Hotelling.

El método de Hotelling basado en el estadístico T^2 es el procedimiento de control más usado a la hora de analizar varias características de forma simultánea.

Hotelling realizó su propuesta basándose en el estadístico T^2

$$T^2 = n(\bar{x} - \mu)'W^{-1}(\bar{x} - \mu)$$

Siendo n el número de observaciones a estudiar y W la covarianza muestral.

El gráfico de Hotelling es considerado una ampliación para el caso multivariante de los gráficos de control de Shewhart.

Los límites de control para este gráfico vienen dados por:

$$\text{LSC} = \frac{p(m-1)(n-1)}{mn-m-p+1} F_{\alpha,p,mn-m-p+1}$$

$$\text{LIC} = 0$$

Siendo p el número de variables a estudio, y m el tamaño muestral tomado en el proceso.

Para poder usar este estadístico, se representa el valor de T^2 frente al número de muestras, y esto es posible a través del paquete `MQSC` del programa estadístico R, ver [16].

En este paquete se incluyen por defecto las funciones que son usadas para obtener el gráfico de control de Hotelling.

Para la comparativa, se usan como los valores absolutos de una distribución de Cauchy, los de una distribución exponencial y de una distribución normal fijando un tamaño muestral de 2 y un nivel de significación de 0.05 para todos los casos.



8.1. Comparativa a través de datos procedentes de distribuciones normales

Los datos usados para la comparativa de los dos gráficos han sido obtenidos a través de dos distribuciones normales de datos, obteniéndose 1000 datos históricos y 500 datos de previsiones.

En el primer gráfico se observan 8 muestras que exceden el valor límite de alarma, pero esto no quiere decir que el proceso se muestre fuera de control, ya que usando el nivel de significación de 0.05, como en este caso, se esperan 12 o 13 muestras fuera de control, por esto este proceso es estable, y como vemos la región del zonoide se ajusta bien a los datos.

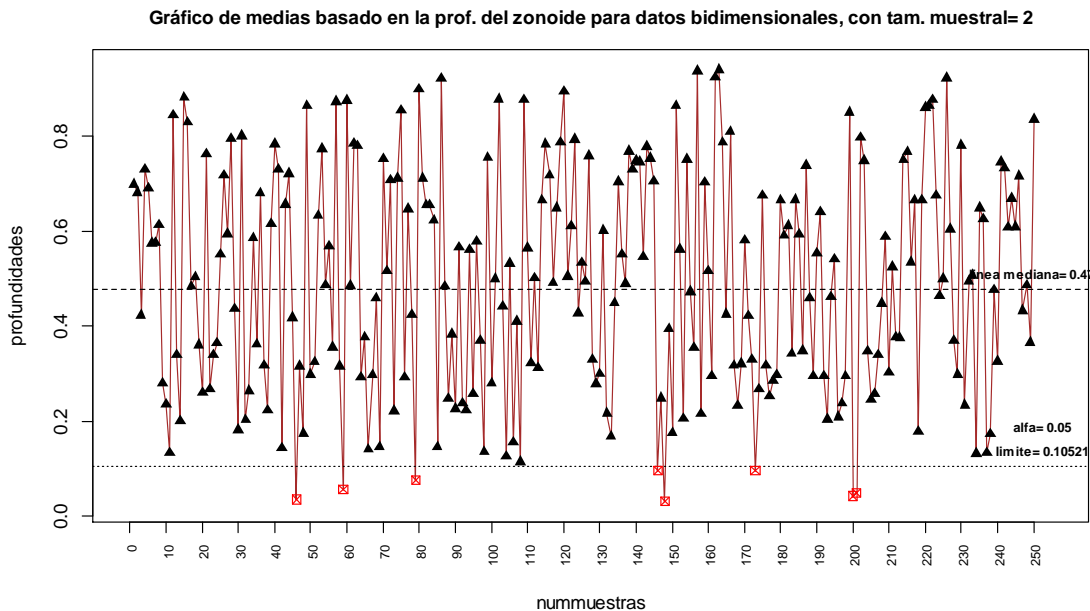


FIGURA 27: "GRÁFICO DE CONTROL BASADO EN LA PROFUNDIDAD DEL ZONOIDE, DATOS NORMALES"

Función en R: `graficobivariantes(matrizpruebas, matrizdatos, 2, 0.05)`

El gráfico de Hotelling para estos mismos datos, procedentes de una normal tiene desempeño parecido en el análisis de datos normales que el grafico de control basado en la profundidad del zonoide, sin embargo aunque detecta el mismo número de anomalías, 7 por 8 anteriormente, es más eficiente el del zonoide ya que la región que se crea para evaluar la estabilidad o no del proyecto, se ajusta mejor a los datos que en el caso del gráfico de control basado en el método de Hotelling.



Hotelling Control Chart

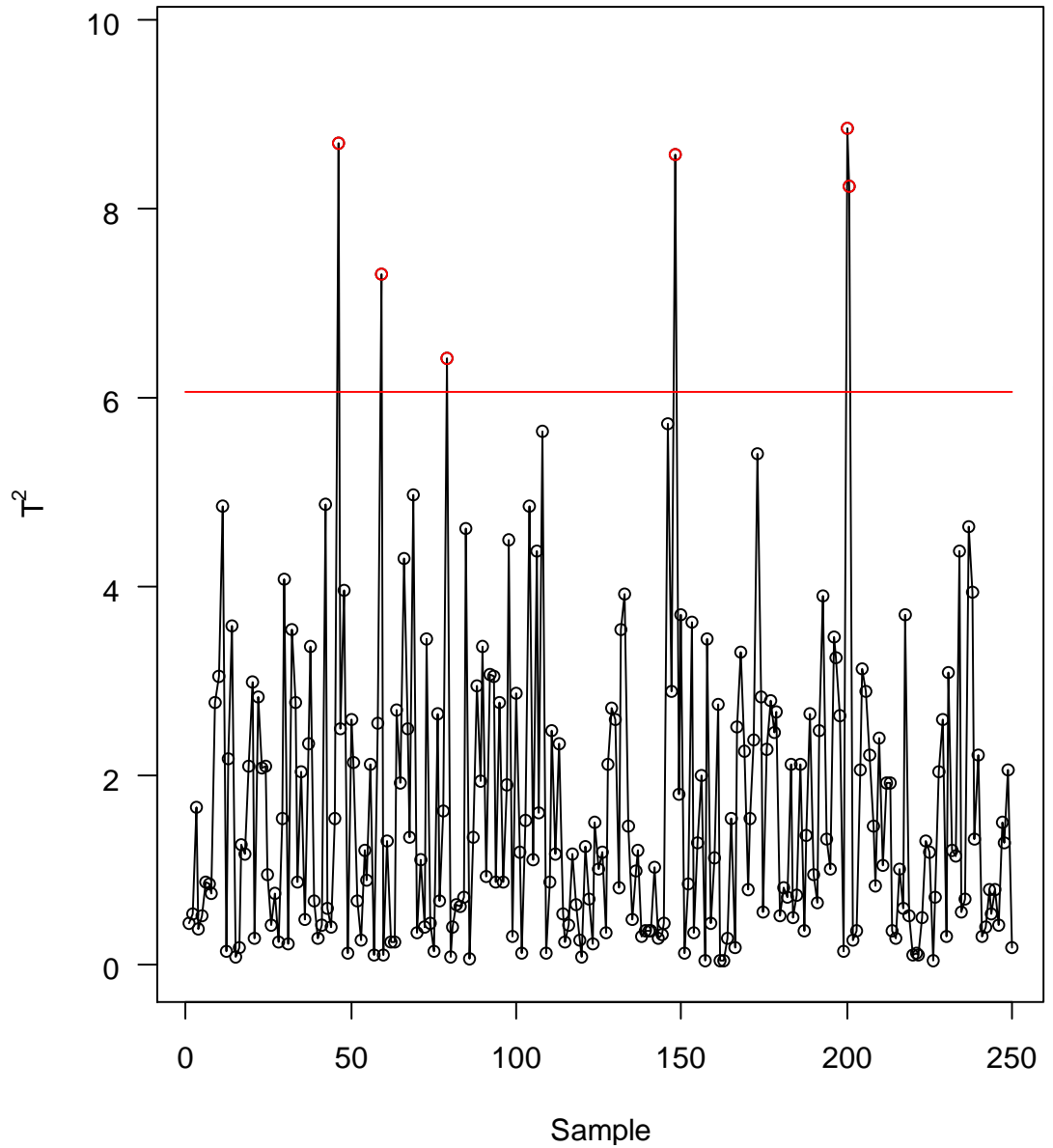


FIGURA 28: "GRÁFICO DE CONTROL BASADO EN EL T^2 DE HOTELLING, DATOS NORMALES"

Función en R: `mult.chart (pruebasMSQC, type = "t2", Xmv = Xmv, S = S, colm = colm)`

Al final se podrá ver una comparativa entre ambos tipos de gráficos.



8.2. Comparativa a través de datos procedentes de distribuciones exponenciales

Los datos usados para la comparativa de los dos gráficos han sido obtenidos a través de las mismas distribuciones exponenciales de datos obteniéndose 1000 datos históricos y 500 datos de previsiones.

A la vista del gráfico de control basado en la profundidad del zonoide, se observa que el proceso se encuentra fuera de control ya que 17 muestras se encuentran sobrepasando el límite de alarma y es esperable para que se encuentre bajo control que sólo 12 o 13 de cada 250 muestras se encuentren fuera de los límites de significación escogido.

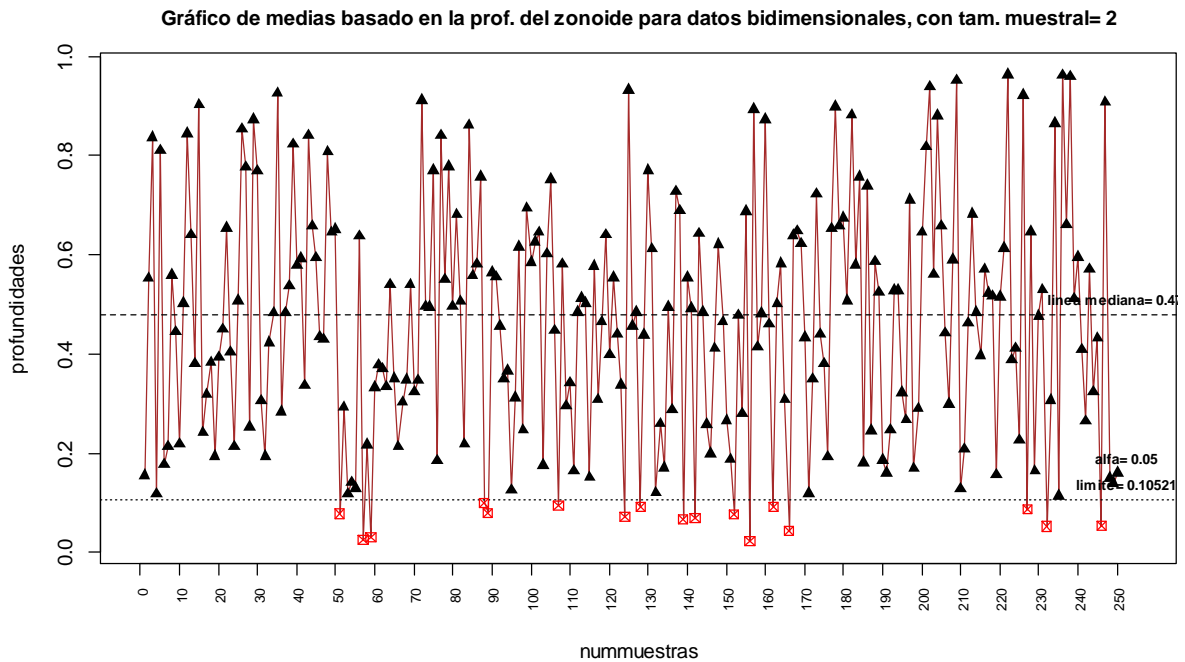


FIGURA 29: "GRÁFICO DE CONTROL BASADO EN LA PROFUNDIDAD DEL ZONOIDE, DATOS EXPONENCIALES"

Función en R: `graficobivariantes(matrizpruebas, matrizdatos, 2, 0.05)`



Hotelling Control Chart

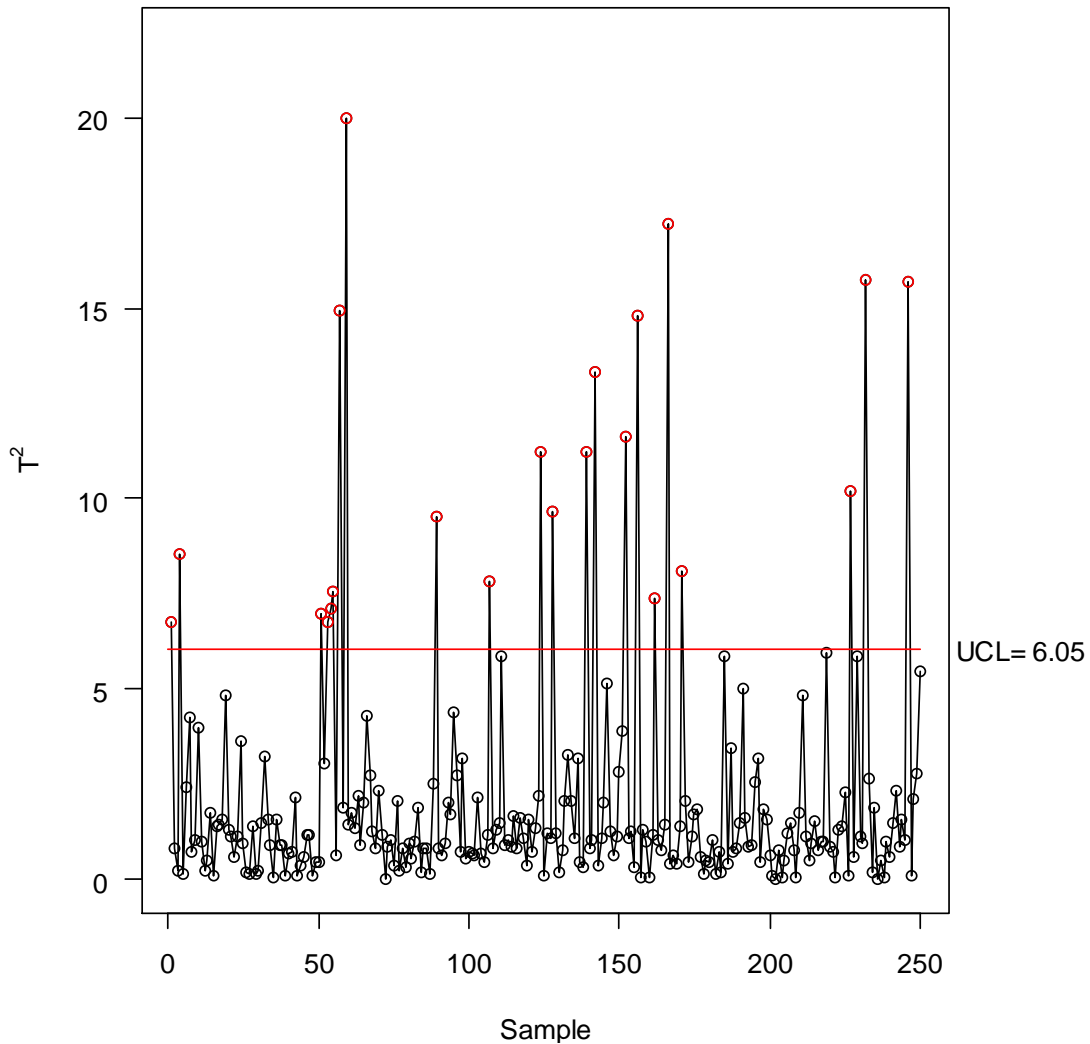


FIGURA 30: "GRÁFICO DE CONTROL BASADO EN EL T^2 DE HOTELLING, DATOS EXPONENCIALES"

Función en R: `mult.chart(pruebasMSQC, type = "t2", Xmv = Xmv, S = S, colm = colm)`

En cuanto al gráfico de control basado en el método de control de Hotelling, se observa que 22 de las 250 muestras analizadas se encuentran por debajo del límite de alarma, por tanto en este análisis el proceso no se encuentra bajo control, y se puede afirmar que el mejor desempeño para realizar el análisis de datos exponenciales es el gráfico de control basado en la profundidad del zonoide por la misma razón que anteriormente.

Este aumento en el número de muestras anómalas que aparecen en el gráfico de control de Hotelling con respecto al de la profundidad del zonoide es debido a la región que crea Hotelling, que al ser elíptica, se ajusta peor a los datos que la región creada por el método de la profundidad del zonoide, que es circular.



8.3. Comparativa a través de datos procedentes de distribuciones de Cauchy

Los datos usados para la comparativa de los dos gráficos han sido obtenidos a través de las mismas distribuciones exponenciales de datos obteniéndose 1000 datos históricos y 500 datos de previsiones.

Se muestran a continuación el histograma de los datos históricos. Se ha realizado a partir de la transformación logarítmica de los datos con el objetivo de que se aprecie lo pesadas que son las colas.

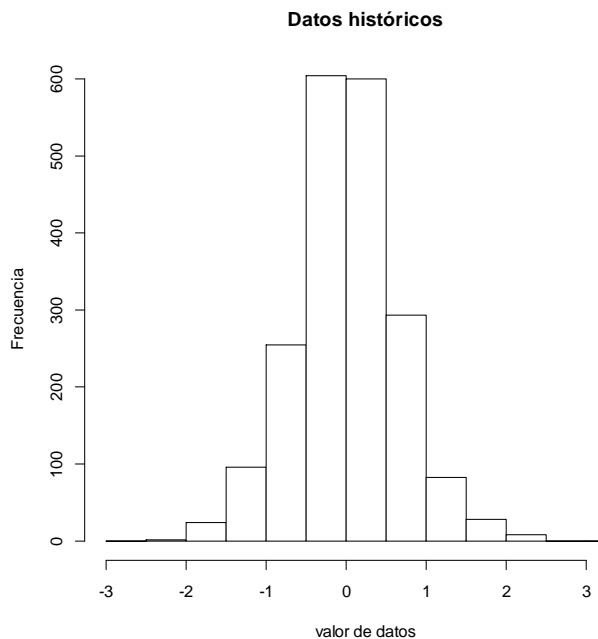


FIGURA 31: "DATOS HISTÓRICOS EN DOS DIMENSIONES"

Se proceden a realizar los gráficos de la media basados en la profundidad del zonoide y en el método T^2 de Hotelling a través de las funciones citadas en la Tabla 8 del Anexo II. Como anteriormente, hay que tener instalado tanto el paquete MSQC y el `dda1pha` del entorno R.

El primero es el gráfico de la profundidad del zonoide.

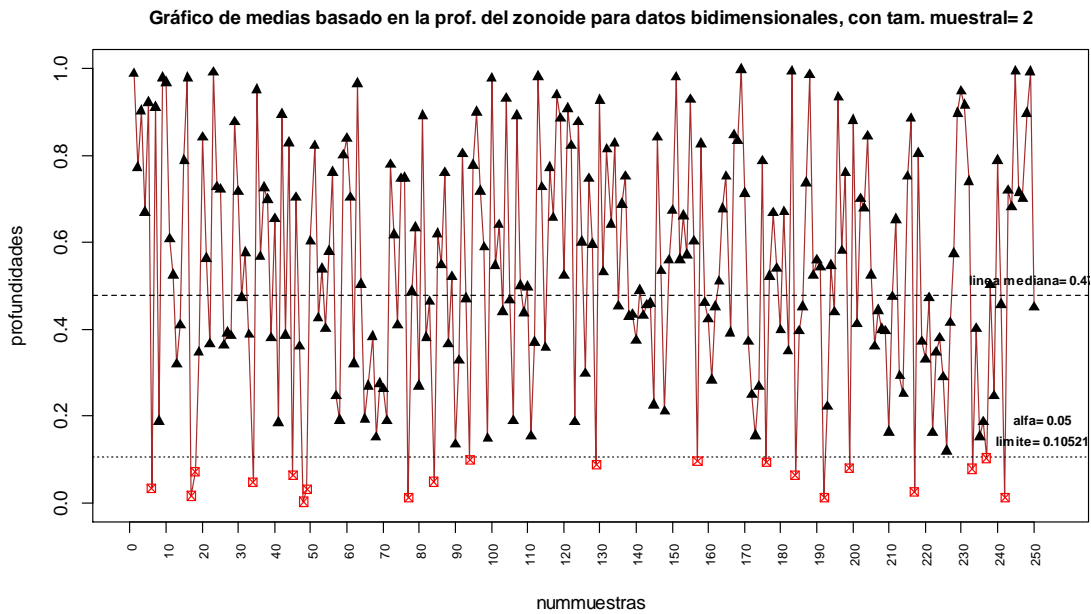


FIGURA 32: "GRÁFICO DE CONTROL BASADO EN LA PROFUNDIDAD DEL ZONOIDE, DATOS CAUCHY"

Función en R: `graficobivariantes(matrizpruebas, matrizdatos, 2, 0.05)`

Se aprecian en el eje de abscisas un total de 250 muestras, de las cuales 20 sobrepasan el límite de alarma del gráfico, es esperable que 5 de cada 100 muestras, se encuentren fuera de control. Aplicando esta proporción, para 250 muestras, es esperable que estén fuera de control 12 o 13 muestras, pero no 17, por tanto para este tipo de gráfico, este proceso no se encuentra bajo control estadístico.

En el gráfico siguiente se muestra el gráfico de control basado en el método de Hotelling para los mismos datos que han sido analizados con el gráfico basado en la profundidad del zonoide.



Hotelling Control Chart

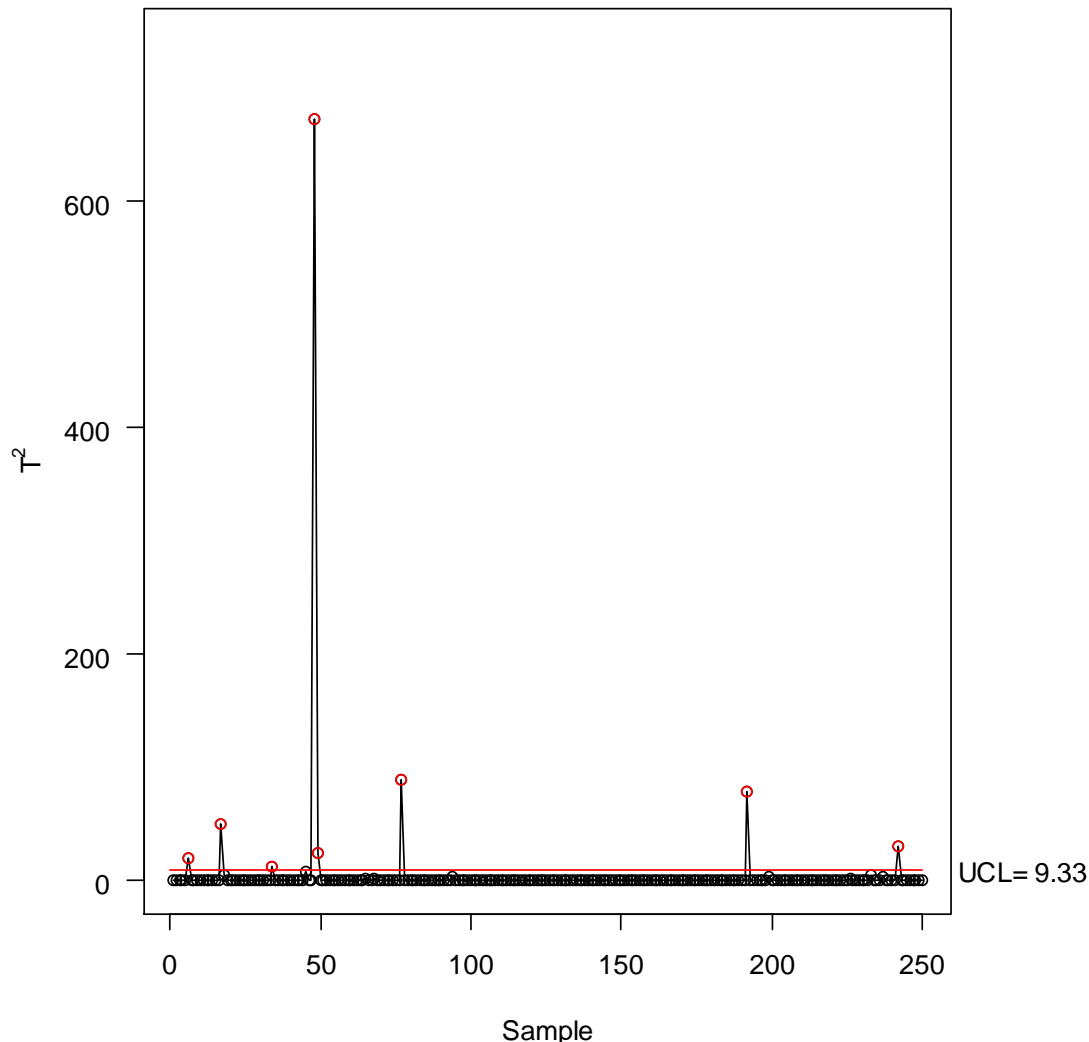


FIGURA 33: "GRÁFICO DE CONTROL DE MEDIAS BASADO EN EL MÉTODO DE HOTELLING"

Función en R: `mult.chart(pruebasMSQC, type = "t2", Xmv = Xmv, S = S, colm = colm)`

Este gráfico indica que solo 8 muestras están fuera de control, pero como se ha visto en el gráfico anterior, basado en la profundidad del zonoide, el proceso no está bajo control, mientras que en el caso del gráfico de Hotelling sí que lo está.

A la vista de estos resultados, se puede decir que para estudiar el caso de distribuciones no normales multidimensionales es más eficiente el gráfico de control de medias basado en la profundidad del zonoide frente al gráfico de Hotelling.

El desempeño a la hora de evaluar el proceso es más efectivo con el método de la profundidad del zonoide debido a que este tipo de gráfico se adapta mejor a los datos, ya que crea una región más centrada que el método de Hotelling que ante



cualquier pequeña variación en la media de las muestras, estas se salen de los límites y no se puede asegurar la estabilidad del proceso.

Por otro lado a la hora de analizar muestras procedentes de distribuciones normales, también es más efectivo el gráfico de control basado en la profundidad del zonoide que el basado en el estadístico T^2 de Hotelling como se ha visto anteriormente, por la misma razón que en el párrafo anterior.



9. Presupuesto

Para el cálculo del presupuesto del proyecto hay que tener en cuenta una serie de factores, tales como las horas de estudio previo, la búsqueda de información, el desarrollo de las funciones en R, la redacción de la memoria, la corrección de ésta, así como el coste del programa estadístico.

Como he dicho al principio del proyecto, el coste del software es cero, ya que es gratuito y de libre distribución, por eso no se reflejan los costes de éste.

El coste total se ha calculado asumiendo un salario medio por trabajador de 30 €/h.

Actividad	Horas dedicadas	Precio/hora	Coste total
Búsqueda y estudio de información	25 h.	30 €/h.	750 €
Desarrollo de los programas	60 h.	30 €/h.	1800 €
Redacción de la memoria	55 h.	30 €/h.	1650 €
Solución de errores	25 h.	30 €/h.	750 €
<u>Total</u>	165 h.	30 €/h.	4950.00 €

El presupuesto para la realización del proyecto es de 4950.00 €, ya que como hemos mencionado anteriormente, el entorno R es gratuito y de libre distribución.



10. Conclusiones

Durante el desarrollo del proyecto han sido expuestos una serie de gráficos y realizado el estudio comparativo entre todos ellos para diferentes tipos de distribuciones de datos, a través del desarrollo de una serie de funciones creadas a partir del programa estadístico R.

Se quería resolver un gran problema a la hora de la elección de los gráficos usados para el análisis de datos; éste es, la suposición de que todos los datos a analizar tenían que pertenecer a una distribución normal y ser unidimensionales, como ocurría en el caso del gráfico de control de Shewhart.

Pero se ha podido comprobar que para el caso de los gráficos basados en la profundidad del zonoide, basados en la profundidad de localización y escala o los de Hotelling, la condición citada anteriormente no es necesaria, ya que son efectivos a la hora de analizar muestras de datos procedentes de distribuciones no normales y en múltiples dimensiones.

En el caso unidimensional se han analizado los gráficos de control de Shewhart, los gráficos de control basados en la profundidad del zonoide y los basados en la profundidad de localización y escala, tanto para datos procedentes de distribuciones normales, como no normales.

Observando los resultados obtenidos, se puede admitir que los gráficos de control de Shewhart contemplan el uso de datos que no sigan una distribución normal en los mismos además de que analizan de una manera fiable datos unidimensionales, al igual que los gráficos basados en la profundidad del zonoide que pueden detectar cambios en la profundidad de las muestras de datos independientemente del tipo de distribución que sigan los datos, obteniendo una medida más eficaz de la posible estabilidad o no del proceso que la que pueden proporcionar otro tipo de gráficos de control.

Mientras que los gráficos basados en la profundidad de localización y escala para datos unidimensionales procedentes de distribuciones no normales muestran anomalías del proceso cuando realmente éstas no existen.

Por tanto, los gráficos de control más fiables a la hora de analizar datos univariantes procedentes de cualquier distribución son los del zonoide ya que las regiones que crea se ajustan mejor a los datos que los demás, que no tienen tanto en cuenta la variabilidad de la media y de la desviación típica de cada una de las muestras.



En el caso multivariante, los gráficos de control de Shewhart no están contemplados, ya que no admiten el uso de distribuciones de datos multidimensionales.

Los gráficos usados en este caso son los del zonoide y lo de Hotelling. Ambos se construyen utilizando el concepto de profundidad estadística y presentan una respuesta parecida cuando los datos pertenecen a una distribución normal, mientras que si la distribución de la que provienen los datos es distinta a una normal, el gráfico basado en el estadístico T^2 de Hotelling no devuelve un análisis fiable, mientras que para el caso del gráfico del zonoide, el uso de unas muestras no normales, no es un problema, ya que este tipo de gráfico no requiere el supuesto de normalidad de las muestras.

Se llega a la conclusión final de que los gráficos más útiles o efectivos a la hora de analizar cualquier conjunto de datos, ya sean multidimensionales o unidimensionales, son los gráficos de control basado en la profundidad del zonoide.

Todos estos gráficos citados anteriormente además de las ventajas anteriores, son muy fácilmente interpretables a la hora de realizar el análisis de la calidad de cualquier proceso industrial.



11. Bibliografía

- [1] MONTGOMERY, D. C. *Control estadístico de la calidad*, 3ª ed., Wiley, Nueva York.
- [2] MONTEAGUDO, José P. y GAITÁN R., Oscar G. (2005). *Herramientas para la gestión energética empresarial*. En *Sciencia et technica* 29 Año XI, pp. 169 - 174.
- [3] R DEVELOPMENT CORE TEAM. *Introducción a R*. Versión 1.0.1. Mayo de 2000. Disponible en web: <https://cran.r-project.org/doc/contrib/R-intro-1.1.0-espanol.1.pdf>
- [4] RUIZ-FALCO, A. *Control estadístico de procesos*. Madrid. Marzo de 2006, pp.: 13-15. Disponible en web
- [5] SHEWHART., W.A. (1931). *The Economic Control of Quality of Manufactured Products*. New York. Western Electric Corporation
- [6] SHAPIRO, S. S. y WILK, M. B. (1965). An analysis-of-variance test for normality. *Biometrika*, 52, pp.: 591-611.
- [7] DEPARTAMENTO DE ESTADÍSTICA UC3M. *Tema 2. Contraste de hipótesis en una población*. En: *Estadística II*. Octubre de 2013. Disponible en web:
http://www.est.uc3m.es/esp/nueva_docencia/getafe/economia/estadistica_ii/documentacion_transp_archivos/tema2esp.pdf
- [8] TUKEY, J.W. (1977). *Exploratory Data Analysis*. Addison-Wesley, Reading, MA.
- [9] LIU, R. (1990). *On a notion of data depth based on random simplices*. *Ann. Statist.*, 18. Pp.: 405-414
- [10] MAHALANOBIS, P. C. (1937). *Normalisation of statistical variates and the use of rectangular coordinates in the theory of sampling distributions*. *Shankya*, 3. Pp.: 35-55.
- [11] DYCKERHOFF, R., KOSHEVOY, G. y MOSLER, K. (1996). *Zonoid data depth: theory and computation*. In: *Prat A. (ed) COMPSTAT 1996 Proceedings in computational statistics*, Physica-Verlag, Heidelberg, pp: 235--240
- [12] MOSLER, K. (2002). *Multivariate Dispersion, Central Regions and Depth: The Lift Zonoid*. Springer, New York
- [13] LÓPEZ, A. (2010). *Similaridad y contraste mediante profundidad estadística*. Tesis doctoral.



- [14] SCRULLA, Luca. *Quality control charts*. Versión 2.6. Octubre de 2014. Disponible en web: <https://cran.r-project.org/web/packages/qcc/qcc.pdf>
- [15] POKOTYLO Oleksii, MOZHAROVSKIY, Pavlo y DYCKERHOFF, Rainer. *Depth-Based Classification and Calculation of Data Depth*. Version 1.1.3.1. Julio de 2015. Disponible en web: <https://cran.r-project.org/web/packages/ddalpha/ddalpha.pdf>
- [16] SANTOS F., Edgar. *Multivariate Statistical Quality Control*. Versión: 1.0.1. Abril de 2014. Disponible en web: <https://cran.r-project.org/web/packages/MSQC/MSQC.pdf>
- [17] FRIAS D., María. D., FERNANDEZ, Jesús y SORDO, Carmen. M. *Tema 7. Control estadística de calidad*. En *Estadística*. Septiembre 2013. Disponible en web: http://ocw.unican.es/enseñanzas-tecnicas/estadistica/materiales-de-clase-1/tema_07.pdf
- [18] EPPSTEIN, D. *Discussion of "Location-Scale Depth" by I. Mizera and C. H. Muller*. Mayo de 2004. Disponible en: <https://www.ics.uci.edu/~eppstein/pubs/Epp-JASA-04.pdf>
- [19] ALVAREZ B., J. *Control estadístico de procesos*. Septiembre de 2011. PP: 29-30, 33, 64-70. Disponible en: http://www.repositoriodigital.ipn.mx/bitstream/handle/123456789/5507/MANUAL_CURSO_CONTROLESTADISTICODEPROCESOS.pdf?sequence=1
- [20] MIZERA, I. y MÜLLER, C. H. (2004). *Location-scale depth*. En *Journal of the American Statistical Association* 99. Pp: 949—989.
- [21] CASCOS, I. (2007). *The expected convex hull trimmed regions of a sample*. En *Computational Statistics* 22, pp: 557—569.
- [22] ZUO, Y. y SERFLING, R. (2000). *General notions of statistical depth function*. *Annals of Statistics* 28, pp: 461—482.
- [23] CASCOS, I. y LÓPEZ-DÍAZ, M. (2012). *Trimmed regions induced by parameters of a probability*. En *Journal of Multivariate Analysis*, 107, pp: 306-318.
- [25] CASCOS, I., LÓPEZ, A. y ROMO, J. *Data depth in Multivariate Statistics*. *Boletín de estadística e Investigación operativa*. Vol. 27, No. 3, pp: 151-170



Anexos

En los anexos se muestran diferentes tablas, entre las cuales se encuentran: los valores de las diferentes constantes para los gráficos de control de Shewhart, los valores de éstas para la profundidad del zonoide y para la profundidad de localización y escala. También se encuentra la tabla que refleja la descripción de las diferentes funciones en R usadas a lo largo del proyecto, así como el código de algunas de las más importantes.

En el último anexo se muestran los datos usados para crear algunos de los gráficos usados en los análisis, como los de Shewhart o los del zonoide para datos unidimensionales.



Anexo I: Constantes para diferentes gráficos de control.

n	A	A2	A3	B3	B4	B5	B6	c4	d2	d3	D1	D2	D3	D4
2	2,121	1,88	2,659	0	3,27	0	2,61	0,8	1,13	0,85	0	3,69	0	3,27
3	1,732	1,02	1,954	0	2,57	0	2,28	0,89	1,69	0,89	0	4,36	0	2,58
4	1,5	0,73	1,628	0	2,27	0	2,09	0,92	2,06	0,88	0	4,7	0	2,28
5	1,342	0,58	1,427	0	2,09	0	1,96	0,94	2,33	0,86	0	4,92	0	2,11
6	1,225	0,48	1,287	0,03	1,97	0,03	1,87	0,95	2,53	0,85	0	5,08	0	2
7	1,134	0,42	1,182	0,12	1,88	0,11	1,81	0,96	2,7	0,83	0,21	5,2	0,08	1,92
8	1,061	0,37	1,099	0,19	1,82	0,18	1,75	0,97	2,85	0,82	0,39	5,31	0,14	1,86
9	1	0,34	1,032	0,24	1,76	0,23	1,71	0,97	2,97	0,81	0,55	5,39	0,18	1,82
10	0,949	0,31	0,975	0,28	1,72	0,28	1,67	0,97	3,08	0,8	0,69	5,47	0,22	1,78
11	0,905	0,29	0,927	0,32	1,68	0,31	1,64	0,98	3,17	0,79	0,81	5,35	0,26	1,74
12	0,866	0,27	0,886	0,35	1,65	0,35	1,61	0,98	3,26	0,78	0,92	5,59	0,28	1,72
13	0,832	0,25	0,85	0,38	1,62	0,37	1,59	0,98	3,34	0,77	1,03	5,65	0,31	1,69
14	0,802	0,24	0,817	0,41	1,59	0,4	1,56	0,98	3,41	0,76	1,12	5,7	0,33	1,67
15	0,775	0,22	0,789	0,43	1,57	0,42	1,54	0,98	3,47	0,76	1,2	5,74	0,35	1,65
16	0,75	0,21	0,763	0,45	1,55	0,44	1,53	0,98	3,53	0,75	1,28	5,78	0,36	1,64
17	0,73	0,2	0,74	0,47	1,53	0,46	1,51	0,99	3,59	0,74	1,36	5,82	0,38	1,62
18	0,707	0,19	0,718	0,48	1,52	0,48	1,5	0,99	3,64	0,74	1,42	5,86	0,39	1,61
19	0,688	0,19	0,698	0,5	1,5	0,79	1,48	0,99	3,69	0,73	1,49	5,89	0,4	1,6
20	0,671	0,18	0,68	0,51	1,49	0,5	1,47	0,99	3,74	0,73	1,55	5,92	0,42	1,59
21	0,655	0,17	0,663	0,52	1,48	0,52	1,46	0,99	3,78	0,72	1,61	5,95	0,43	1,58
22	0,64	0,17	0,647	0,53	1,47	0,53	1,45	0,99	3,82	0,72	1,66	5,98	0,44	1,57
23	0,626	0,16	0,633	0,55	1,46	0,54	1,38	0,99	3,86	0,72	1,71	6,01	0,44	1,56
24	0,612	0,16	0,619	0,56	1,45	0,55	1,43	0,99	3,9	0,71	1,76	6,03	0,44	1,55
25	0,6	0,15	0,606	0,57	1,44	0,56	1,42	0,99	3,93	0,71	1,81	6,06	0,46	1,54

TABLA 3: "CONSTANTES PARA GRÁFICOS DE CONTROL DE SHEWHART"



dimensión = 1

<i>k / α</i>	<i>0,5</i>	<i>0,05</i>	<i>0,01</i>	<i>0,0027</i>
<i>1</i>	0,57965	0,06363	0,01297	0,353
<i>2</i>	0,71355	0,220503	0,08674	0,04339
<i>3</i>	0,77335	0,31313	0,17044	0,10494
<i>4</i>	0,80851	0,39211	0,24302	0,16639
<i>5</i>	0,83214	0,45182	0,30334	0,22163
<i>6</i>	0,84932	0,49858	0,3535	0,26995
<i>7</i>	0,86249	0,5363	0,39568	0,31201
<i>8</i>	0,87297	0,56746	0,43161	0,34877
<i>9</i>	0,88156	0,59371	0,46259	0,38109
<i>10</i>	0,88875	0,61618	0,4896	0,40971
<i>11</i>	0,89488	0,63566	0,51339	0,43522
<i>12</i>	0,90018	0,65276	0,53451	0,45812
<i>13</i>	0,90481	0,6679	0,55341	0,4788
<i>14</i>	0,90891	0,68142	0,57045	0,49756
<i>15</i>	0,91257	0,69359	0,58589	0,51469

TABLA 4: "CONSTANTES PARA GRÁFICOS DE CONTROL BASADOS EN LA PROFUNDIDAD DEL ZONOIDE, DATOS EN 1 DIMENSIÓN"

dimensión = 2

<i>k / α</i>	<i>0,5</i>	<i>0,05</i>	<i>0,01</i>	<i>0,0027</i>
<i>1</i>	0,29137	0,01858	0,00315	0,00077
<i>2</i>	0,47848	0,10521	0,04084	0,0194
<i>3</i>	0,57615	0,19524	0,10059	0,05996
<i>4</i>	0,63719	0,27033	0,161	0,10768
<i>5</i>	0,67955	0,33138	0,21568	0,15477
<i>6</i>	0,71099	0,38147	0,26371	0,19847
<i>7</i>	0,73542	0,4232	0,30566	0,23811
<i>8</i>	0,75507	0,45849	0,3424	0,27382
<i>9</i>	0,77128	0,48876	0,37477	0,30596
<i>10</i>	0,78492	0,51503	0,40346	0,33495
<i>11</i>	0,79661	0,53809	0,42907	0,36117
<i>12</i>	0,80674	0,55851	0,54208	0,385
<i>13</i>	0,81564	0,57674	0,47287	0,40674
<i>14</i>	0,82353	0,59313	0,49175	0,42665
<i>15</i>	0,83058	0,60797	0,50899	0,44496

TABLA 5: "CONSTANTES PARA GRÁFICOS DE CONTROL BASADOS EN LA PROFUNDIDAD DEL ZONOIDE, DATOS EN 2 DIMENSIONES"



dimensión = 3

<i>k/α</i>	0,5	0,05	0,01	0,0027
1	0,15475	0,675	0,001	0,00022
2	0,33493	0,06122	0,02224	0,01014
3	0,4494	0,13349	0,0659	0,03826
4	0,51818	0,20074	0,11588	0,07604
5	0,57079	0,25887	0,16442	0,11624
6	0,61069	0,30845	0,20901	0,15539
7	0,64216	0,35088	0,24918	0,19213
8	0,66774	0,38748	0,28518	0,22606
9	0,68902	0,41935	0,31744	0,25717
10	0,70706	0,44735	0,34645	0,28565
11	0,72257	0,47216	0,37264	0,31175
12	0,7361	0,49341	0,39638	0,33569
13	0,74801	0,51422	0,418	0,35773
14	0,7586	0,53223	0,43777	0,37805
15	0,76809	0,5486	0,45593	0,39686

TABLA 6: "CONSTANTES PARA GRÁFICOS DE CONTROL BASADOS EN LA PROFUNDIDAD DEL ZONOIDE, DATOS EN 3 DIMENSIONES"

<i>k/α</i>	0,5	0,05	0,01	0,027
2	0,36133	0,0395	0,00825	0,0224
3	0,51893	0,15084	0,06929	0,03804
4	0,59666	0,23063	0,12974	0,0837
5	0,64565	0,28716	0,17721	0,12251
6	0,68042	0,32877	0,21471	0,15625
7	0,70534	0,36089	0,24597	0,18052

TABLA 7: "CONSTANTES PARA LOS GRÁFICOS DE CONTROL BASADOS EN LA PROFUNDIDAD DE LOCALIZACIÓN Y ESCALA"



Anexo II: Lógica de las funciones usadas durante la realización del proyecto en la implementación de los gráficos de control unidimensionales y multidimensionales.

Función	Descripción
<p><i>graficorangos(datos,n,alfa)</i></p>	<p style="text-align: center;">Entrada</p> <p><i>Datos:</i> Vector de muestras de la característica a analizar, en este caso, el rango. <i>n:</i> Tamaño muestral. <i>alfa:</i> Nivel de significación escogido.</p> <p style="text-align: center;">Salida</p> <p>Gráfico de control de Shewhart para los rangos en la fase I.</p>
<p><i>graficomedias(datos,n,alfa)</i></p>	<p style="text-align: center;">Entrada</p> <p><i>Datos:</i> Vector de muestras de la característica a analizar, en este caso, la media. <i>n:</i> Tamaño muestral. <i>alfa:</i> Nivel de significación escogido.</p> <p style="text-align: center;">Salida</p> <p>Gráfico de control de Shewhart para la media en la fase I.</p>
<p><i>controlrangos2</i> <i>(matrizdatos,n,LCS,lineacentral,LCI)</i></p>	<p style="text-align: center;">Entrada</p> <p><i>matrizdatos:</i> Matriz de muestras de la característica a analizar, en este caso, el rango. <i>n:</i> Tamaño muestral. <i>LCS:</i> Límite superior de control obtenido en la fase I. <i>Línea central:</i> Valor al que se aproxima el rango en la fase I. <i>LCI:</i> Límite inferior de control de la fase I.</p> <p style="text-align: center;">Salida</p> <p>Gráfico de control de Shewhart para el rango de la fase II.</p>



<p><i>controldesvestandar(datos,n,alfa)</i></p>	<p style="text-align: center;">Entrada</p> <p><i>datos</i>: Vector de observaciones de la desviación estándar de los datos. <i>n</i>: Tamaño muestral. <i>alfa</i>: Nivel de significación.</p> <p style="text-align: center;">Salida</p> <p>Gráfico de control de Shewhart para la desviación estándar del proceso.</p>
<p><i>controlmedias(datos,n,alfa)</i></p>	<p style="text-align: center;">Entrada</p> <p><i>datos</i>: Vector de observaciones de la media del proceso. <i>n</i>: Tamaño muestral. <i>alfa</i>: Nivel de significación.</p> <p style="text-align: center;">Salida</p> <p>Gráfico de control de Shewhart para la media del proceso.</p>
<p><i>hist(datos, main="Muestras de datos históricos")</i></p>	<p style="text-align: center;">Entrada</p> <p><i>Datos</i>: Datos históricos usados para el análisis o datos de prueba.</p> <p style="text-align: center;">Salida</p> <p>Histograma de los datos históricos usados para el gráfico de la profundidad del zonoide con datos unidimensionales.</p>
<p><i>mediasmuestrales(matriz,k)</i></p>	<p style="text-align: center;">Entrada</p> <p><i>matriz</i>: Matriz de las observaciones de los datos de prueba usados para el análisis tanto para la profundidad del zonoide como para la profundidad de localización y escala. <i>k</i>: Tamaño muestral escogido.</p> <p style="text-align: center;">Salida</p> <p>Valores de las medias de cada una de las muestras escogidas.</p>



<p><i>zdepth (s,x,ord=F)</i></p>	<p style="text-align: center;">Entrada</p> <p><i>s</i>: Medias de las muestras de datos usadas para el análisis. <i>x</i>: Matriz de datos históricos usados para el análisis a través del gráfico de control basado en la profundidad del zonoide. <i>Ord</i>: Puede ser T o F, dependiendo de si la matriz de datos es ordenada o no, en este caso es F.</p> <p style="text-align: center;">Salida</p> <p>Valores de las profundidades de cada una de las muestras tomadas a través de los datos de prueba.</p>
<p><i>graficounidimensional (x,datoshistoricos,n,alfa)</i></p>	<p style="text-align: center;">Entrada</p> <p><i>x</i>: matriz de datos de prueba <i>datoshistoricos</i>: matriz de datos históricos sobre los que se calcula la profundidad del zonoide. <i>n</i>: Tamaño muestral elegido. <i>alfa</i>: Nivel de significación.</p> <p style="text-align: center;">Salida</p> <p>Gráfico de control de la media basado en la profundidad del zonoide, para datos unidimensionales.</p>
<p><i>graficobivariantes (matrizp,matrizd,k,alfa)</i></p>	<p style="text-align: center;">Entrada</p> <p><i>Matrizp</i>: Matriz de datos de prueba, siendo los datos bidimensionales. <i>Matrizd</i>: Matriz de datos históricos, siendo datos bidimensionales. <i>k</i>: Tamaño muestral. <i>alfa</i>: Nivel de significación escogido.</p> <p style="text-align: center;">Salida</p> <p>Gráfico de control de la media basado en la profundidad del zonoide para datos bidimensionales.</p>



<p><i>graficotrivariantes</i> (<i>matrizp,matrizd,k,alfa</i>)</p>	<p style="text-align: center;">Entrada</p> <p><i>Matrizp</i>: Matriz de datos de prueba, siendo los datos tridimensionales. <i>Matrizd</i>: Matriz de datos históricos, siendo datos tridimensionales. <i>k</i>: Tamaño muestral. <i>alfa</i>: Nivel de significación escogido.</p> <p style="text-align: center;">Salida</p> <p>Gráfico de control de la media basado en la profundidad del zonoide para datos tridimensionales.</p>
<p><i>Graficolocalizacionescala</i> (<i>matrizp,matrizd,k,alfa</i>)</p>	<p style="text-align: center;">Entrada</p> <p><i>Matrizp</i>: Matriz de datos de prueba, siendo los datos bidimensionales. <i>Matrizd</i>: Matriz de datos históricos, siendo datos bidimensionales. <i>k</i>: Tamaño muestral. <i>alfa</i>: Nivel de significación escogido.</p> <p style="text-align: center;">Salida</p> <p>Gráfico de control de la media basado en la profundidad de localización y escala, siendo los datos bidimensionales.</p>
<p><i>mult.chart(type = "t2", datosMSQC,alpha=0.05)</i></p>	<p style="text-align: center;">Entrada</p> <p><i>type</i>: Tipo de gráfico que se va a realizar. <i>datosMSQC</i>: Datos a usar en el análisis para la obtención de límites de control confiables. <i>alpha</i>: Nivel de significación escogido, en este caso 0.05.</p> <p style="text-align: center;">Salida</p> <p>Gráfico de control de la primera fase de Hotelling, donde se calculan un límite de control fiable para el estudio de los datos.</p>



<p><i>mult.chart(pruebasMSQC, type = "t2", Xmv = Xmv, S = S, colm = colm, alpha = 0.05)</i></p>	<p style="text-align: center;">Entrada</p> <p><i>type</i>: Tipo de gráfico que se va a realizar. <i>pruebasMSQC</i>: Datos de previsiones que se quiere analizar. <i>alpha</i>: Nivel de significación escogido, en este caso 0.05. <i>Xmv</i>: Es el vector de medias. <i>colm</i>: Número de muestras de datos.</p> <p style="text-align: center;">Salida</p> <p>Gráfico de control de la segunda fase de Hotelling, donde se procede a señalar si existen anomalías en un proceso.</p>
---	---

TABLA 8: "FUNCIONES CREADAS EN R QUE APARECEN DURANTE EL PROYECTO Y DEFINICIÓN DE SUS PRINCIPALES ARGUMENTOS"



Anexo III: Código de las funciones más representativas para la obtención de los gráficos del proyecto.

<i>Zdepth(s,x,ord=F)</i>	<pre> # Función de profundidad del zonoide para datos unidimensionales. zdepth<-function(s,x,ord=F){ if(ord==F){x=sort(x)} n=length(x) i=1 if((s<x[1]) (s>x[n])){zdepth=0;return(zdepth)} if(s==x[1]){while(x[i+1]==s){i=i+1};zdepth=i/n; return(zdepth)} if(s==x[n]){while(x[n-i]==s) {i=i+1}; zdepth=i/n; return(zdepth)} m=mean(x) if(s==m){zdepth=1;return(zdepth)} if(s<m){ suma=x[1] while(((suma+x[i+1])/(i+1)<=s)&(i<n)){i=i+1; suma=suma+x[i]} zdepth=i*(suma/i-x[i+1])/(n*(s-x[i+1])) return(zdepth) } if(s>m){ suma=x[n] while(((suma+x[n-i])/(i+1)>=s) &(i<n)) {i=i+1; suma=suma+x[n-i+1]} zdepth=i*(suma/i-x[n-i])/(n*(s-x[n-i])) return(zdepth) } } </pre>
--------------------------	---



Graficounidimensional(x,datoshistoricos,n,alfa)

```

# Cálculo del gráfico de control de la media de datos
# unidimensionales utilizando la profundidad del zonoide.

# Datos y pruebas en el anexo IV.
datos
pruebas
shapiro.test(datos)

# Función que calcula la profundidad del zonoide con datos
unidimensionales.
vzdepth=function(s,x,ord=F){
  k=length(s)
  vzdepth=vector(length=k)
  x=sort(x)
  for(i in 1:k){
    vzdepth[i]=zdepth(s[i],x,T)
  }
  return(vzdepth)
}

# Una vez que los datos pertenecen a una normal, se realiza el
# gráfico de control.

graficounidimensional <- function(x,datoshistoricos,n,alfa){

  x<- matrix(x,nrow=n)

# Colocamos los límites en función del valor alfa y del tamaño
# muestral, que aparecen en la tabla 4 del anexo I.
# Para alfa = 0.0027

  if(alfa==0.0027 && nrow(x)==2) {limite<- 0.04339}
  if(alfa==0.0027 && nrow(x)==3) {limite<- 0.10494}
  if(alfa==0.0027 && nrow(x)==4) {limite<- 0.16639}
  if(alfa==0.0027 && nrow(x)==5) {limite<- 0.22163}
  if(alfa==0.0027 && nrow(x)==6) {limite<- 0.26995}
  if(alfa==0.0027 && nrow(x)==7) {limite<- 0.31201}
  if(alfa==0.0027 && nrow(x)==8) {limite<- 0.34877}
  if(alfa==0.0027 && nrow(x)==9) {limite<- 0.38109}
  if(alfa==0.0027 && nrow(x)==10) {limite<- 0.40971}

#Para alfa=0.01

  if(alfa==0.01 && nrow(x)==2) {limite<- 0.08674}
  if(alfa==0.01 && nrow(x)==3) {limite<- 0.17044}
  if(alfa==0.01 && nrow(x)==4) {limite<- 0.24302}
  if(alfa==0.01 && nrow(x)==5) {limite<- 0.30334}
  if(alfa==0.01 && nrow(x)==6) {limite<- 0.3535}
  if(alfa==0.01 && nrow(x)==7) {limite<- 0.39568}
  if(alfa==0.01 && nrow(x)==8) {limite<- 0.43161}
  if(alfa==0.01 && nrow(x)==9) {limite<- 0.46259}
  if(alfa==0.01 && nrow(x)==10) {limite<- 0.4896}

# Para alfa=0.05

  if(alfa==0.05 && nrow(x)==2) {limite<- 0.20503}
  if(alfa==0.05 && nrow(x)==3) {limite<- 0.31313}
  if(alfa==0.05 && nrow(x)==4) {limite<- 0.39211}

```



Graficounidimensional(x,datoshistoricos,n,alfa)

```

if(alfa==0.05 && nrow(x)==5) {limite<- 0.45182}
if(alfa==0.05 && nrow(x)==6) {limite<- 0.49858}
if(alfa==0.05 && nrow(x)==7) {limite<- 0.53630}
if(alfa==0.05 && nrow(x)==8) {limite<- 0.56746}
if(alfa==0.05 && nrow(x)==9) {limite<- 0.59371}
if(alfa==0.05 && nrow(x)==10) {limite<- 0.61618}

# Línea central para un valor de alfa=0.5

if(nrow(x)==2){lineamediana<-0.71355}
if(nrow(x)==3){lineamediana<-0.77335}
if(nrow(x)==4){lineamediana<-0.80851}
if(nrow(x)==5){lineamediana<-0.83214}
if(nrow(x)==6){lineamediana<-0.84932}
if(nrow(x)==7){lineamediana<-0.86249}
if(nrow(x)==8){lineamediana<-0.87297}
if(nrow(x)==9){lineamediana<-0.88156}
if(nrow(x)==10){lineamediana<-0.89488}

# Calculamos la media de cada una de las muestras y aplicamos la
# función de profundidad a las medias calculadas con respecto de
# los datos históricos.
mediasx <- c(apply(x,2,mean))

# Llamamos a la función que nos calcula las profundidades del
# zonoide de los datos de prueba frente a los datos históricos.
# Se almacena en la variable profundidades.

profundidades<-vzdepth(mediasx, datoshistoricos)

# Una vez calculadas las profundidades se procede a graficar.

nummuestras<-1:length(profundidades)

dev.new(width=10, height=6)

plot(nummuestras, profundidades, type="l", xaxt="n",
      col="brown", xlim=c(0,(length(profundidades)+5)),
      ylim=c(min(limite,min(profundidades))-0.005,
max(profundidades)+0.005))

axis(1, 1:length(profundidades), cex.axis=0.7, las=2)
abline(h=limite, lty=3)
text((max(nummuestras)),limite, paste("limite=",limite), n,
pos=3, font=2, cex=0.7)
abline(h=lineamediana, lty=2)
text((max(nummuestras)),lineamediana, paste("linea
mediana=",lineamediana), pos=3,font=2, cex=0.7)
text((max(nummuestras)), limite+0.03,
paste("alfa=",alfa),pos=3,font=2, cex=0.7)

for(i in 1:length(profundidades)){

if(profundidades[i]<limite){

```



<i>graficounidimensional(x,datos,historicos,n,alfa)</i>	<pre> points(nummuestras[i], profundidades[i], pch=7, col="red") } else{ points(nummuestras[i], profundidades[i], col="black", pch=17) } } mtext(paste("Gráfico de medias basado en la profundidad del zonoide con datos unidimensionales con tamaño muestral=",n),side=3, font=2,line=1) } # Cogemos algunos tamaños muestrales y niveles de significación # distintos con el objetivo de analizar en diferentes escenarios. graficounidimensional(pruebas, datos, 4, 0.0027) graficounidimensional(pruebas, datos, 2, 0.01) graficounidimensional(pruebas, datos, 5, 0.05) </pre>
<i>graficobivariantes(matrizp,matrizd,k,alfa)</i>	<pre> # Gráfico de medias de datos bidimensionales de prueba frente a los # datos históricos basado en la profundidad del zonaide # Para poder usar las funciones de cálculo se debe instalar el # paquete dalpha. # Datos procedentes de una normal de media 35 y desviación típica # 0.5, siendo x los datos históricos e y los de la previsión. x<-rnorm(400,35,0.5) x matrizdatos<-matrix(x, ncol=2, byrow=FALSE) matrizdatos y<- rnorm(100,35,0.5) y matrizpruebas<- matrix(y,ncol=2, byrow=FALSE) matrizpruebas # Mostramos a través de un histograma los datos históricos plot(matrizdatos, main="Datos históricos") mediasmuestrales<-function(matriz,k){ n<-nrow(matriz) media<-matrix(ncol=2,nrow=(n/k)) for(i in 1:(n/k)){ media[i,1]=mean(matriz[(1+((i-1)*k)):i*k],1]) media[i,2]=mean(matriz[(1+((i-1)*k)):i*k],2]) } media } </pre>



graficobivariantes(matrizp,matrizd,k,alfa)

```
# Una vez calculadas las medias de esta matriz de pruebas, pasamos
# a hacer el gráfico basado en la profundidad del zonoide de
# nuestros datos de prueba frente a los históricos.

graficobivariantes<-function(matrizp,matrizd,k,alfa){

# Límites de control procedentes de la tabla 5 del Anexo I.
  # Para alfa = 0.0027

  if(alfa==0.0027 && k==2) {limite<- 0.0194}
  if(alfa==0.0027 && k==3) {limite<- 0.05996}
  if(alfa==0.0027 && k==4) {limite<- 0.10768}
  if(alfa==0.0027 && k==5) {limite<- 0.15477}
  if(alfa==0.0027 && k==6) {limite<- 0.19847}
  if(alfa==0.0027 && k==7) {limite<- 0.23811}
  if(alfa==0.0027 && k==8) {limite<- 0.27382}
  if(alfa==0.0027 && k==9) {limite<- 0.30596}
  if(alfa==0.0027 && k==10) {limite<- 0.33495}

  # Para alfa = 0.01

  if(alfa==0.01 && k==2) {limite<- 0.04084}
  if(alfa==0.01 && k==3) {limite<- 0.10059}
  if(alfa==0.01 && k==4) {limite<- 0.161}
  if(alfa==0.01 && k==5) {limite<- 0.21568}
  if(alfa==0.01 && k==6) {limite<- 0.26371}
  if(alfa==0.01 && k==7) {limite<- 0.30566}
  if(alfa==0.01 && k==8) {limite<- 0.3424}
  if(alfa==0.01 && k==9) {limite<- 0.37477}
  if(alfa==0.01 && k==10) {limite<- 0.40346}

  #Para alfa = 0.05

  if(alfa==0.05 && k==2) {limite<- 0.10521}
  if(alfa==0.05 && k==3) {limite<- 0.19524}
  if(alfa==0.05 && k==4) {limite<- 0.27033}
  if(alfa==0.05 && k==5) {limite<- 0.33138}
  if(alfa==0.05 && k==6) {limite<- 0.38147}
  if(alfa==0.05 && k==7) {limite<- 0.4232}
  if(alfa==0.05 && k==8) {limite<- 0.45849}
  if(alfa==0.05 && k==9) {limite<- 0.48876}
  if(alfa==0.05 && k==10) {limite<- 0.51503}

  # Línea central para alfa = 0.5

  if(k==2){lineamediana<- 0.47848}
  if(k==3){lineamediana<- 0.57615}
  if(k==4){lineamediana<- 0.63719}
  if(k==5){lineamediana<- 0.67955}
  if(k==6){lineamediana<- 0.71099}
  if(k==7){lineamediana<- 0.73542}
  if(k==8){lineamediana<- 0.75507}
  if(k==9){lineamediana<- 0.77128}
  if(k==10){lineamediana<- 0.78492}
```



Graficobivariantes(matrizp,matrizd,k,alfa)

```

medias<-mediasmuestrales(matrizp,k)

# Se llama a la función que proporciona el paquete dalpha para el
# cálculo de la profundidad del zonoide.

profundidades<-depth.zonoid(medias,matrizd)
nummuestras<-1:length(profundidades)

dev.new(width=10, height=6)

# Se realiza el gráfico colocando cada muestra con su profundidad.

plot(nummuestras, profundidades, type="l", xaxt="n", col="brown",
xlim=c(0, (length(profundidades)+5)),
      ylim=c(min(limite,min(profundidades))-0.005,
max(profundidades)+0.005))

  axis(1, 1:length(profundidades), cex.axis=0.7, las=2)
  abline(h=limite, lty=3)
  text((max(nummuestras)+2.5),limite,
paste("limite=",limite),pos=3, font=2, cex=0.7)

  abline(h=lineamediana, lty=2)
  text((max(nummuestras)+2.5),lineamediana, paste("linea
mediana=",lineamediana), pos=3,
      font=2, cex=0.7)

  text((max(nummuestras)+2.5), limite+0.05,
paste("alfa=",alfa),k,pos=3,font=2, cex=0.7)

  for(i in 1:length(profundidades)){

    if(profundidades[i]<limite){
      points(nummuestras[i], profundidades[i], pch=7, col="red")
    }
    else{
      points(nummuestras[i], profundidades[i], col="black", pch=17)
    }
  }

  mtext(paste("Gráfico de medias basado en la profundidad del
zonoide para datos bidimensionales, con tamaño muestral=",k),
side=3, font=2,line=1)

}

# Se evalúan tres tamaños muestrales y tres niveles de
# significación con el fin de evaluar el proceso en diferentes
# escenarios.

graficobivariantes(matrizpruebas, matrizdatos, 4, 0.0027)
graficobivariantes(matrizpruebas, matrizdatos, 2, 0.01)
graficobivariantes(matrizpruebas, matrizdatos, 2, 0.05)

```



Gráfico localización y escala (matrizp, matrizd, k, alfa)

```
# Gráfico que muestra la profundidad de localización y escala de
# una serie de muestras de datos bidimensionales previstos frente
# a una serie de datos históricos del proceso.

# Se realiza el ejemplo para unos datos bidimensionales procedentes
# de una normal con 200 datos históricos y 50 de previsiones
# futuras.

x<-rnorm(400, 35,0.5)
x
matrizdatos<-matrix(c(x,x^2), ncol=2, byrow=FALSE)
matrizdatos

y<- rnorm(100,35,0.5)
y
matrizpruebas<- matrix(c(y,y^2),ncol=2, byrow=FALSE)
matrizpruebas

# Función que calcula la media de cada una de las muestras de los
# datos de prueba

mediasmuestraes<-function(matriz,k){
  n<-nrow(matriz)
  media<-matrix(ncol=2,nrow=(n/k))

  for(i in 1:(n/k)){
    media[i,1]=mean(matriz[(1+((i-1)*k):(i*k),1])
    media[i,2]=mean(matriz[(1+((i-1)*k):(i*k),2])
  }
  media
}

graficolocalizacionescalea<-function(matrizp,matrizd,k,alfa){

# Se colocan los límites en función del valor alfa y del tamaño
# muestral procedentes de la tabla 7 del Anexo I.

# Para alfa = 0.0027

if(alfa==0.0027 && k==2) {limite<- 0.00224}
if(alfa==0.0027 && k==3) {limite<- 0.03804}
if(alfa==0.0027 && k==4) {limite<- 0.08370}
if(alfa==0.0027 && k==5) {limite<- 0.12251}
if(alfa==0.0027 && k==6) {limite<- 0.15625}
if(alfa==0.0027 && k==7) {limite<- 0.18052}

#Para alfa = 0.01

if(alfa==0.01 && k==2) {limite<- 0.00825}
if(alfa==0.01 && k==3) {limite<- 0.06929}
if(alfa==0.01 && k==4) {limite<- 0.12974}
if(alfa==0.01 && k==5) {limite<- 0.17721}
if(alfa==0.01 && k==6) {limite<- 0.21471}
if(alfa==0.01 && k==7) {limite<- 0.24597}
```




Graficocalizacionescala(matrizp,matrizd,k,alfa)

```
# Para alfa = 0.05

if(alfa==0.05 && k==2) {limite<- 0.03950}
if(alfa==0.05 && k==3) {limite<- 0.15084}
if(alfa==0.05 && k==4) {limite<- 0.23063}
if(alfa==0.05 && k==5) {limite<- 0.28716}
if(alfa==0.05 && k==6) {limite<- 0.32887}
if(alfa==0.05 && k==7) {limite<- 0.36089}

# línea central para un valor alfa=0.5

if(k==2){lineamediana<- 0.36133}
if(k==3){lineamediana<- 0.51893}
if(k==4){lineamediana<- 0.59666}
if(k==5){lineamediana<- 0.64565}
if(k==6){lineamediana<- 0.68042}
if(k==7){lineamediana<- 0.70534}

# Llamamos a mediasmuestrales para calcular las medias de las
# muestras de prueba, y se calculan las profundidades de éstas
# frente a los datos históricos.

medias<-mediasmuestrales(matrizp,k)

profundidades<-depth.zonoid(medias,matrizd)

# Se comienza a graficar.

nummuestras<-1:length(profundidades)

dev.new(width=10, height=6)
plot(nummuestras, profundidades, type="l", xaxt="n", col="brown",
xlim=c(0, (length(profundidades)+5)),
ylim=c(min(limite,min(profundidades))-0.005,
max(profundidades)+0.005))

axis(1, 1:length(profundidades), cex.axis=0.7, las=2)

abline(h=limite, lty=3)

text((max(nummuestras)+2.5),limite, paste("limite=",limite),
pos=3, font=2, cex=0.7)

abline(h=lineamediana, lty=2)

text((max(nummuestras)+2.5),lineamediana, paste("linea mediana=",
lineamediana), pos=3,
font=2, cex=0.7)
```



<p><i>Graficolocalizacionescale(matrizp,matrizd,k,alfa)</i></p>	<pre> text((max(nummuestras)+2.5), limite+0.05, paste("alfa=",alfa),k,pos=3,font=2, cex=0.7) for(i in 1:length(profundidades)){ if(profundidades[i]<limite){ points(nummuestras[i], profundidades[i], pch=7, col="red") } else{ points(nummuestras[i], profundidades[i], col="black", pch=17) } } mtext(paste("Gráfico de localización y escala con tamaño muestral=",k), side=3, font=2,line=1) } Gráfico para tres niveles de significación y tamaños muestrales diferentes. graficolocalizacionescale(matrizpruebas, matrizdatos, 5, 0.0027) graficolocalizacionescale(matrizpruebas, matrizdatos, 2, 0.01) graficolocalizacionescale(matrizpruebas, matrizdatos, 2, 0.05) </pre>
---	---

TABLA 9: "FUNCIONES DESTACADAS EN LA REALIZACIÓN DEL PROYECTO"

Anexo IV: Datos para la realización de los gráficos de control de Shewhart y para los gráficos de control basados en la profundidad del zonoide con datos unidimensionales.

Para los gráficos de control de Shewhart para el rango, la media y la desviación típica.

Datos de la primera fase.

```

datos1<- c(34.999, 34.998, 35.085, 34.986, 35.073, 34.958, 34.901,
34.958, 35.055, 35.002, 35.007, 35.037, 35.037, 34.945, 35.053,
35.049, 34.996, 35.008, 34.965, 35.055, 34.949, 35.018, 34.958,
34.957, 35.036, 35.053, 34.933, 35.011, 35.021, 35.020, 35.031,
34.958, 35.078, 35.064, 34.992, 35.029, 35.064, 35.007, 34.988,
35.071, 35.040, 34.956, 35.089, 35.047, 35.055, 35.055, 35.007,
34.935, 34.998, 34.998, 34.979, 34.921, 35.000, 34.977, 34.982,
35.036, 35.032, 34.984, 35.077, 34.998, 35.016, 35.102, 35.033,
34.980, 35.009, 34.896, 35.004, 34.940, 34.961, 34.985, 34.987,
35.025, 35.012, 34.984, 34.944, 34.981, 34.915, 34.997, 35.044,
35.052, 34.977, 34.914, 34.891, 35.022, 34.996, 34.986, 35.068,
35.083, 34.868, 34.947, 34.984, 35.086, 35.088, 34.970, 35.016,

```



34.964, 35.091, 34.974, 35.033, 35.007, 35.067, 35.010, 35.001,
34.997, 34.925, 35.065, 35.014, 35.006, 35.044, 34.972, 35.032,
34.967, 34.993, 35.002, 35.045, 34.948, 34.986, 34.993, 35.036,
34.938, 34.938, 35.094, 35.019, 34.998, 35.078, 35.034, 34.980,
35.001)

Datos de la segunda fase.

```
datos2<- c(34.958, 35.020, 35.032, 35.012, 35.044, 35.087, 34.929,  
34.998, 35.025, 35.035, 35.052, 34.938, 35.032, 34.935, 34.923,  
35.032, 34.936, 34.903, 35.046, 34.966, 34.953, 35.043, 35.001,  
34.932, 35.051, 35.060, 34.962, 35.014, 35.049, 35.073, 35.023,  
35.053, 34.976, 34.990, 35.042, 35.023, 35.057, 34.975, 35.012,  
34.980, 34.945, 34.964, 35.026, 35.000, 35.067, 35.026, 34.966,  
34.979, 34.978, 35.013, 34.991, 34.945, 35.017, 34.997, 35.017,  
34.992, 34.965, 35.062, 34.974, 34.977, 34.867, 35.039, 35.030,  
35.026, 35.000, 34.942, 35.059, 35.010, 34.977, 35.033, 34.921,  
34.980, 34.966, 35.004, 35.055, 34.919, 34.969, 34.958, 34.979,  
34.971, 35.026, 35.017, 35.028, 35.034, 35.033, 35.028, 34.964,  
34.967, 35.011, 35.019, 34.926, 35.045, 34.944, 34.994, 34.988,  
35.081, 35.088, 35.005, 35.061, 34.957)
```

Para el gráfico de control basado en la profundidad del zonoide con datos normales unidimensionales.

Datos históricos.

```
datos1<- c(34.999, 34.998, 35.085, 34.986, 35.073, 34.958, 34.901,  
34.958, 35.055, 35.002, 35.007, 35.037, 35.037, 34.945, 35.053,  
35.049, 34.996, 35.008, 34.965, 35.055, 34.949, 34.980, 35.001,  
35.018, 34.958, 34.957, 35.036, 35.053, 34.933, 35.011, 35.021,  
35.020, 35.031, 34.958, 35.078, 35.064, 34.992, 35.029, 35.064,  
35.007, 34.988, 35.071, 35.040, 34.956, 35.089, 35.047, 35.055,  
35.055, 35.007, 34.935, 34.998, 34.998, 34.979, 34.921, 35.000,  
34.977, 34.982, 35.036, 35.032, 34.984, 35.077, 34.998, 35.016,  
35.102, 35.033, 34.980, 35.009, 34.896, 35.004, 34.940, 34.961,  
34.985, 34.987, 35.025, 35.012, 34.984, 34.944, 34.981, 34.915,  
34.997, 35.044, 35.052, 34.977, 34.914, 34.891, 35.022, 34.996,  
34.986, 35.068, 35.083, 34.868, 34.947, 34.984, 35.086, 35.088,  
34.970, 35.016, 34.964, 35.091, 34.974, 35.033, 35.007, 35.067,  
35.010, 35.001, 34.997, 34.925, 35.065, 35.014, 35.006, 35.044,  
34.972, 35.032, 34.967, 34.993, 35.002, 35.045, 34.948, 34.986,  
34.993, 35.036, 34.938, 34.938, 35.094, 35.019, 34.998, 35.078,  
35.034)
```

Datos previstos.

```
pruebas<- c(34.972, 35.006, 35.165, 35.019, 34.976, 34.970, 35.017,  
34.997, 34.976, 35.037, 35.070, 34.996, 34.985, 34.970, 35.096,  
35.040, 35.060, 35.083, 34.993, 34.962, 35.020, 35.036, 35.009,  
34.945, 35.027, 34.932, 35.131, 35.021, 34.996, 35.009, 35.007,  
34.990, 35.077, 35.101, 34.982, 34.880, 34.954, 34.945, 34.952,  
35.019, 35.014, 34.996, 34.926, 34.952, 34.941, 34.945, 34.919,  
35.064, 35.032, 34.934, 35.074, 34.977, 35.076, 35.013, 34.943,
```



35.022, 34.988, 34.959, 34.979, 35.027, 34.954, 34.956, 34.925,
35.127, 34.939, 34.939, 34.978, 35.067, 35.052, 34.868, 35.037,
35.050, 35.011, 34.972, 35.000, 35.035, 35.005, 34.994, 34.980,
35.037, 34.976, 34.943, 34.972, 34.984, 35.062, 34.941, 35.009,
34.975, 35.033, 35.051, 35.094, 34.990)