



**INSTITUTO TECNOLÓGICO
SUPERIOR DE GUANAJUATO**

**“CONTROL ESTADÍSTICO MULTIVARIADO
DE PROCESOS BASADO EN COMPONENTES
PRINCIPALES PARA DATOS DEL SECTOR
INDUSTRIAL”**

TESIS

Presenta:

Juana Edith Lozano Hernández

Con la asesoría de:

**M.A. María Guadalupe de Lourdes
Acosta Castillo**



**INSTITUTO TECNOLÓGICO
SUPERIOR DE GUANAJUATO**

**“Control estadístico multivariado de procesos basado en
componentes principales para datos del sector industrial”**

Tesis

Que para obtener el grado de:

INGENIERO INDUSTRIAL

Presenta:

Juana Edith Lozano Hernández

Con la asesoría de:

Mtra. María Guadalupe de Lourdes Acosta Castillo

Y la Co-asesoría de:

Mtro. Jorge Luis Morales Martínez

Guanajuato, Gto. 07 de junio de 2019

Agradecimientos

Agradezco a Dios, por permitirme concluir satisfactoriamente esta etapa de mi vida y por siempre llenarme de bendiciones .

A mis padres porque gracias a su trabajo, educación y a los valores que me inculcaron, sigo en el camino de la vida profesional.

A mis hermanos por ser mi fuente inspiración y motivación.

A mi novio por su compañía y apoyo incondicional en los momentos más difíciles durante la lucha por concluir la universidad.

A Inés y Toquina por sus palabras de aliento y motivación, así como todo su apoyo tanto económico como emocional.

A mis asesores de tesis por su valioso apoyo para la elaboración de esta tesis.

A la Coordinadora de Ing. Industrial Nicté-Ha Mejía, la Coordinadora de Ing. en Sistemas Computacionales Beatriz Cisneros y a todos mis profesores de la carrera, los que me impartieron alguna clase y los que no también, pues cada uno apporto distintas enseñanzas y oportunidades para mi formación.

Resumen

El objetivo de la presente investigación es mostrar una aplicación del control estadístico multivariado de procesos (MSPC) sobre un conjunto reducido de variables. Complementando con la determinación de la capacidad del proceso. Se analizaron dos conjuntos de datos con 28 mediciones de cinco características de calidad del proceso de producción de bimetales para termostatos. En análisis se realizó en dos etapas: Fase I y II, para cada una, se estudió el comportamiento univariado y multivariado de las variables. Luego, se aplicó un análisis de componente principales (PCA) para reducir la dimensión del problema. Posteriormente se utilizó la técnica de cartas de control T2 de Hotelling basada en PCA. Finalmente, se determinaron, los índices multivariados de capacidad MC_p , MC_{pk} y MC_{pm} . Con el análisis de normalidad univariada y multivariada, se observó que la no normalidad para una de las variables en la Fase II, no afectó para que el conjunto de las cinco variables siguiera una distribución multinormal. Por otro lado, con la aplicación del PCA, resultó ser suficiente trabajar con solo dos componentes principales las cuales representaron el 80.61% y el 83.85% de la variabilidad, para la Fase I y II respectivamente. Finalmente, con la implementación de las cartas de control T2 de Hotelling se observó que el proceso en la Fase I estuvo en control, mientras que en la Fase II se detectó un punto fuera del límite de control, indicando la presencia de causas especiales y por lo tanto de un proceso fuera de control.

Índice general

Agradecimientos.....	3
Resumen.....	4
Índice general.....	5
Índice de tablas.....	7
Índice de figuras.....	8
Introducción.....	9
Capítulo 1. Generalidades del proyecto.....	10
1.1. Planteamiento del problema.....	10
1.2. Justificación.....	10
1.3. Objetivos.....	11
1.3.1. Objetivo general.....	11
1.3.2. Objetivos específicos.....	12
Capítulo 2. Marco teórico.....	13
2.1. Fundamento teórico.....	13
2.1.1. Industria en México.....	13
2.1.2. Conceptos básicos de calidad.....	16
2.1.3. Análisis de Componentes Principales.....	19
2.1.4. Índices de capacidad de procesos univariados y multivariados.....	21
2.1.5. Control estadístico de procesos.....	32
2.1.6. Control estadístico multivariado de procesos.....	40
2.2. Antecedentes.....	46
2.3. El Control de procesos y el cumplimiento de los requisitos de las Normas..	47
Capítulo 3. Metodología.....	49
3.1. Materiales y métodos.....	49
3.2. Descripción del caso.....	49
3.3. Aplicación de los métodos al caso de estudio.....	51
3.3.1. Análisis del estado del proceso en Fase I.....	52
3.3.2. Análisis del estado del proceso en Fase II.....	62
Capítulo 4. Discusión y análisis de resultados.....	69
4.1. Resultados.....	69

4.2. Importancia o trascendencia	70
Conclusiones y recomendaciones.....	70
Anexo	72
Código implementado en el Software R para generar los gráficos y resultados estadísticos.	72
Referencias	78

Índice de tablas

Tabla 1. Estados con mayor producción minera en febrero de 2019.	16
Tabla 2. Patrones no aleatorios que se pueden encontrar en un gráfico de control..	39
Tabla 3. Mediciones por variable para la Fase I.....	50
Tabla 4. Mediciones por variable para la Fase II.....	51
Tabla 5. Especificaciones para las variables.....	51
Tabla 6. Medidas de tendencia central y dispersión para cada una de las variables analizadas.	53
Tabla 7. Resultados de la prueba de Jarque-Bera para normalidad.	55
Tabla 8. Matriz de varianzas-covarianzas de las variables.	57
Tabla 9. Valores propios de la matriz de covarianzas de las variables.	57
Tabla 10. Vectores propios de la matriz de covarianzas de las variables.	57
Tabla 11. Matriz de varianzas para las componentes principales.	57
Tabla 12. Tabla de puntuaciones de las componentes principales.	59
Tabla 13. Medidas de tendencia central y dispersión para cada una de las variables analizadas en la Fase II.	63
Tabla 14. Resultados de la prueba de Jarque-Bera para normalidad en la Fase II...	64
Tabla 15. Matriz de varianzas y covarianzas de las variables en la Fase II.	65
Tabla 16. Matriz de varianzas para las componentes principales de la Fase II.	66

Índice de figuras

Figura 1. Elementos de un proceso de producción.	16
Figura 2. Anchura de las medidas distribuidas normales de una característica del producto (adaptada de (de- Felipe & Benedito, 2017)).	23
Figura 3. Regiones de proceso y especificación ((adaptada de (de- Felipe & Benedito, 2017)).	25
Figura 4. PR y SR modificada según K. Chang et al. (1988) (a), Modificación PR' según L. Pearn et al. (1992) (b).....	27
Figura 5. Representación de los ejes principales de la PR. Pearn et al. (1992) (b). .	28
Figura 6. Descripción de PR con sus vectores propios. Pearn et al. (1992) (b).	30
Figura 7. Tipos de errores en la conclusión de un proceso.	34
Figura 8. Componentes de un gráfico de control. Gutiérrez y De la Vara (2013).	37
Figura 9. Cartas de control bidimensional en comparación con dos cartas de control univariadas (Adaptada de Mason y Young (2002)).	41
Figura 10. Métodos utilizados en el desarrollo la investigación.....	49
Figura 11. Histograma de las variables individuales.	52
Figura 12. Gráficos Q-Q de las variables individuales.....	54
Figura 13. Matriz de dispersión entre las variables.	56
Figura 14. Matriz de dispersión entre las variables.	56
Figura 15. Diagrama de Pareto que muestra el porcentaje de variación de las componentes principales.....	58
Figura 16. Elipse de confianza para los puntos de las dos componentes principales.	60
Figura 17. Gráfico de control T2 de Hotelling para las componentes principales.....	60
Figura 18. Histograma de las variables individuales para la Fase II.....	62
Figura 19. Gráficos Q-Q de las variables individuales para la Fase II.....	63
Figura 20. Matriz de dispersión entre las variables para la Fase II.	64
Figura 21. Matriz de dispersión entre las variables para la Fase II.	65
Figura 22. Diagrama de Pareto que muestra el porcentaje de variación de las componentes principales en la Fase II.	66
Figura 23. Elipse de confianza para las componentes principales de la Fase II.	67
Figura 24. Gráfico de control T2 de Hotelling para las componentes principales de la Fase II.	68

Introducción

Con el crecimiento de la industria manufacturera, la competencia entre las empresas fabricantes ha incrementado. Además, los clientes se han vuelto cada vez más exigentes con la calidad de los productos que adquieren. A causa de esto, las empresas han enfocado sus esfuerzos en el control y mejora de la calidad. Para responder al crecimiento de la demanda, actualmente las empresas del sector industrial han implementado la automatización de los procesos, generando así una producción más acelerada. Así mismo, para garantizar la calidad del producto a resultado necesario medir distintas características de calidad, lo que conlleva a obtener bases de datos muy grandes. Afortunadamente en la actualidad existen distintos métodos de reducción de dimensiones que facilitan el trabajo con conjuntos grandes de variables para aplicar control de procesos. El objetivo de esta tesis es mostrar una aplicación de control estadístico multivariado de procesos para analizar el estado de un proceso, en el cual se miden cinco características de calidad. La aplicación se realizó en dos etapas: Fase I y Fase II. Para cada etapa, realizó el análisis de componentes principales, y posteriormente, sobre las componentes principales generadas se procedió al control estadístico de procesos generando gráficos de control T2 de Hotelling. Para finalizar, se efectuó el estudio de capacidad del proceso mediante la determinación de los índices de capacidad.

Con el propósito de proporcionar las bases teóricas necesarias para la comprensión de esta investigación, en el capítulo 2 se redacta lo referente a los conceptos, herramientas y métodos utilizados, así mismo se incluye previamente una síntesis de la situación actual en la industria. La metodología se desarrolla en el capítulo 3. Al inicio de este capítulo se incluye un diagrama que muestra de manera general la secuencia para la aplicación de los métodos, que posteriormente se desarrollan haciendo uso del software estadístico R. Los resultados obtenidos se resumen en el Capítulo 4. Finalmente, en la última sección se presentan las conclusiones y algunas recomendaciones derivadas de la observación y el análisis de los resultados de la sección anterior.

Capítulo 1. Generalidades del proyecto

1.1. Planteamiento del problema

Debido al incremento de la competitividad en el sector industrial, ha resultado necesario para las empresas manufactureras conocer constantemente el estado de sus procesos. Más aún, las normas para Sistemas de gestión de la calidad indican como requisito que las empresas lleven a cabo el control de sus procesos. Para realizar el control de procesos se recurre a la aplicación de herramientas estadísticas que permitan detectar variaciones y predecir el comportamiento del proceso. Una de las herramientas más utilizadas es el Control Estadístico de Procesos.

El control estadístico univariado de procesos se enfoca en el estudio del comportamiento del proceso evaluando solo una variable. Sin embargo, durante el desarrollo de un proceso influyen varios factores: mano de obra, maquinaria, material, entre otros, que pueden causar variación en distintas etapas del desarrollo del proceso. Por lo cual, se deben estudiar más de una característica de calidad, es decir, se tendrán que medir más de una variable, añadiendo que dichas variables podrían estar fuertemente interrelacionadas. Al aplicar control estadístico univariado se omite la interrelación entre las variables, y los efectos que dicha interrelación puedan generar. Como consecuencia, se ha incurrido en la necesidad de que los procesos univariados se conviertan en procesos multivariados.

1.2. Justificación

En los procesos de producción actuales, para controlar la calidad de los productos es necesario vigilar más de una característica simultáneamente (Anchiraico, 2003). El control estadístico de procesos (Statistical process control, SPC), es una de las herramientas más utilizadas para evaluar el comportamiento de las características de calidad a lo largo del proceso. La técnica de control estadístico de procesos más utilizada es el gráfico de control. Sin embargo debido a que el SPC se basa en estadísticas univariadas y gráficos de control univariados, en los que solo una variable se monitorea y / o visualiza a la vez (Camacho, Pérez, García y Maciá 2016), su implementación para un conjunto de varias variables podría traer consigo una serie de

desventajas, entre ellas: 1. El número de gráficos de control a implementar y vigilar puede llegar a ser relativamente grande, esto dependiendo del número de características que se deseen controlar. 2. Al realizar el control de las características de manera independiente se está obviando cualquier tipo de relación que pudiera existir entre las características de calidad (Mosquera, Olaya y Escobar, 2007). Por lo tanto, a medida que aumenta el número de variables que se monitorean, el enfoque para visualizar una variable a la vez reduce su rendimiento y es más conveniente utilizar estadísticas y gráficos multivariados (Camacho et al., 2016), es decir resulta de mayor utilidad aplicar el Control estadístico multivariado de procesos (Multivariate statistical process control, MSPC). Sin embargo, Kourti y MacGregor afirman que (como se citó en Camacho et al., 2016) el MSPC tradicional no tiene en cuenta la correlación entre las variables, lo que puede llevar a problemas de precisión y computación debido al mal acondicionamiento. Concluyendo entonces, que cuando el número de variables es muy grande y / o las variables están altamente interrelacionadas, se recomienda el uso de métodos de variables latentes, como Análisis de componentes principales (Principal Component Analysis, PCA) (Camacho et al., 2016). En esta tesis se realiza una evaluación completa del estado de un proceso en el sector industrial mediante la aplicación del análisis de componentes principales y posteriormente la aplicación de Control estadístico multivariado de procesos.

Es importante mencionar que en la literatura existen pocas investigaciones sobre la aplicación de componentes principales y el control estadístico multivariado de procesos a conjuntos de datos del sector industrial. Por lo cual la actual investigación aporta nuevos resultados que pueden ser de utilidad para considerarlos como base para iniciar otras investigaciones.

1.3. Objetivos

1.3.1. Objetivo general

Analizar el estado y la capacidad de un proceso de producción aplicando control estadístico multivariado de procesos basado en componentes principales.

1.3.2. Objetivos específicos

- Analizar los datos de las variables de estudio medidas, determinando su tendencia central y dispersión, así como su distribución, mediante el uso de herramientas estadísticas de calidad.
- Analizar si las variables de estudio tienen distribución multinormal, aplicando pruebas de normalidad multivariada a los datos utilizados.
- Aplicar análisis de componentes principales para reducir la dimensión de las matrices de datos.
- Obtener los gráficos de control de los conjuntos de datos multivariados para analizar la estabilidad y variabilidad del proceso correspondiente.
- Determinar los índices de capacidad del proceso.
- Utilizar herramientas de programación del software libre R.

Capítulo 2. Marco teórico

2.1. Fundamento teórico

Se considera importante tener el conocimiento sobre la situación actual en la industria, por ser el sector en el que enfocamos la aplicación de esta tesis. Por lo tanto, en este capítulo iniciamos con una síntesis de los aspectos sobresalientes de los subsectores más importantes del sector industrial en México.

2.1.1. Industria en México

La economía de México está compuesta por diversos sectores económicos, siendo uno de los más importantes el sector industrial, el cual, de acuerdo con datos del Instituto Nacional de Estadística y Geografía (INEGI, 2019) aportó el 28.76% de PIB en el último trimestre de 2018.

México es un país rico en recursos naturales, mineros y de mano de obra, por lo que la industria que se desarrolla en México es muy variada. Las actividades económicas del sector industrial más importantes se describen a continuación.

Industria automotriz

La industria automotriz está conformada por las organizaciones enfocadas en el diseño, manufactura y venta de automóviles, está integrada por dos sectores: terminal y autopartes.

Desde hace varios años han ingresado al territorio mexicano grandes empresas extranjeras, tanto productoras de autopartes como armadoras, produciendo piezas para consumo de México y de otros países. De acuerdo con la Confederación de Asociaciones de Agentes Aduanales de la República Mexicana (CAAAREM), en 2018 México se mantuvo como el cuarto exportador de vehículos a nivel mundial. Las exportaciones fueron principalmente a Estados Unidos con 74.4% del total, seguido por Canadá con 7.2%, Alemania con 4.4%, Brasil con 2.3% y Colombia con 1.5%. Según datos de la revista Forbes publicados en enero de 2019 México sigue manteniéndose en esa posición.

Entre las principales armadoras extranjeras que se han instalado en México se encuentran Land Rover, General Motors, Mazda, Peugeot, Honda, Ford Motor, Nissan, Toyota, Volkswagen, Audi y FCA México.

Las asociaciones que representan a la industria automotriz son:

- Dentro de México: la Asociación Mexicana de la Industria Automotriz (AMIA) establecida el 3 de enero de 1951. AMIA tiene como principal misión dar a conocer cuestiones automotrices, evolución y estadística de los fabricantes de vehículos, afiliados en esta asociación (AMIA, 2011). Las principales empresas que la conforman son: como socios fundadores FCA México, Ford Motor Company, General Motors de México, Nissan Mexicana, Volkswagen de México; y como socios Afiliados Honda de México, Mazda Motor de México, Toyota Motor Sales de México y BMW de México.
- A nivel internacional: la Organización Internacional de Constructores de Automóviles (OICA), que está conformada por 38 asociaciones comerciales nacionales de distintas partes del mundo (OICA, 2019).

Industria Alimentaria

La industria alimentaria comprende la transformación de productos agropecuarios y es una de las que más aportan al PIB de México. De acuerdo con datos publicados por la Secretaría de Economía (SE, 2018), la industria de alimentos procesados representó el 23.4% del PIB manufacturero y 3.9% del PIB total. En 2016, las exportaciones mexicanas de la industria alcanzaron un valor de 8,324 mdd. El principal destino de las exportaciones mexicanas fue Estados Unidos, con 65% de participación, seguido de Venezuela (3%), Sudáfrica y Guatemala (2.5%) respectivamente y Canadá (2.3%). Al igual que otras, la industria alimentaria "está compuesta por nueve subsectores (...):

1. Elaboración de alimentos para animales
2. Molienda de granos y de semillas y obtención de aceites y grasas.
3. Elaboración de azúcares, chocolates, dulces y similares.
4. Conservación de frutas, verduras, guisos y otros alimentos preparados.
5. Elaboración de productos lácteos.
6. Matanza, empaque y procesamiento de carne de ganado, aves y otros animales comestibles.

7. Preparación y envasado de pescados y mariscos.
8. Elaboración de productos de panadería y tortillas.
9. Otras industrias alimentarias” (CIAL, 2018, p. 5).

Algunas de las empresas de la industria alimentaria que se encuentran en México son: Grupo bimbo, grupo alfa, Arca continental, Bachoco y Lala.

Industria minero metalúrgica

La industria minero metalúrgica el 4 % del PIB nacional. Se divide en dos subsectores la industria minera y la industria manufacturera:

“Minería. Se conoce como Minería a toda aquella actividad productiva en la que se extraen, explotan y/o benefician los minerales depositados en el suelo y en el subsuelo.

Manufactura. Las industrias relacionadas con la transformación de minerales se denominan industrias metálicas básicas, las cuales realizan procesos siderúrgicos y metalúrgicos, así como a las industrias que fabrican bienes a base de minerales no metálicos.” (INEGI, 2014, p. 8)

De acuerdo con la estadística de la industria minero metalúrgica publicada por INEGI, los estados con mayor producción minera durante febrero de 2019 son Sonora, Chihuahua, Guerrero, Durango, Zacatecas y San Luis Potosí (ver tabla 1).

México es un país rico en recursos minerales y actualmente se ubica entre los 10 principales productores de: plata, bismuto, fluorita, celestita, wollastonita, cadmio, molibdeno, plomo, zinc, diatomita, sal, barita, grafito, yeso, oro y cobre. Ocupando el primer lugar en producción de Plata a nivel mundial (SE, 2019).

Estados con mayor producción minera		
Mineral	Estado	Toneladas
Oro	Sonora	3203
	Chihuahua	1537
	Guerrero	1513
Plata	Zacatecas	186680
	Chihuahua	90015
	Durango	53906
Plomo	Zacatecas	10819
	Chihuahua	3153
	Durango	1817
Cobre	Sonora	46379
	Zacatecas	3443
	San Luis Potosí	2480
Zinc	Zacatecas	21760
	Chihuahua	8595
	Durango	7767

Tabla 1. Estados con mayor producción minera en febrero de 2019.

2.1.2. Conceptos básicos de calidad

La competencia en el sector industrial es más fuerte cada día, por lo que las empresas deben poner atención en que las expectativas de los clientes sobre los productos que consumen sean cubiertas al máximo. Por lo tanto, la empresa debe atender y mejorar constantemente su proceso de producción, el cual consiste en un conjunto de actividades desarrolladas para modificar una entrada o insumo y que dan como resultado una salida (o producto) (figura 1).

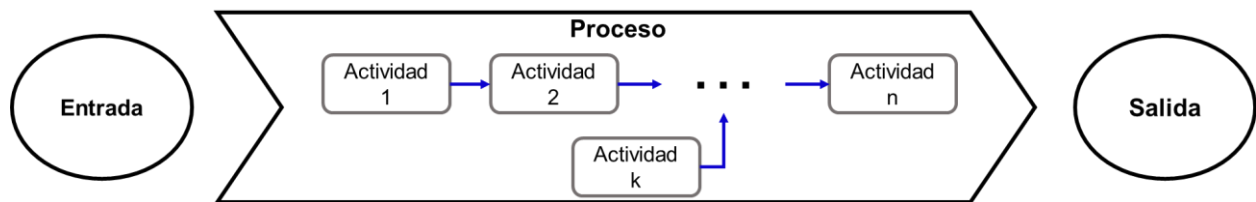


Figura 1. Elementos de un proceso de producción.

Ahora bien, en un proceso se debe asegurar el cumplimiento de la producción planeada y cuidar que ésta posea un nivel alto de calidad.

Muchos autores han definido el término “Calidad” de distintas maneras. López (2006) proporciona una lista de definiciones propuestas por expertos a lo largo de la historia:

- J.M. Juran: “La idoneidad para el uso”.
- Feigenbaum: “El conjunto de las características del producto (bien o servicio) de marketing, ingeniería, fabricación y mantenimiento a través del cual el producto satisface las expectativas del cliente”.
- Crosby: “Cumplimiento de unas especificaciones o la conformidad de unos requisitos”.
- W.E. Deming: “Un grado predecible de uniformidad y habilidad a bajo costo y adecuado a las necesidades del mercado”.

En su caso particular López (2006) afirma que la calidad de una empresa debe desarrollarse en dos direcciones:

- La calidad para el cliente: (...) refiriéndose a Estudiar las necesidades del cliente (...), mejorar las prestaciones (...) y controlar la satisfacción (...). Y,
- La calidad para la empresa: (...) se centra en (...) mejora tecnológica del proceso (...), mejora económica del proceso (...) y supervivencia como objetivo prioritario (...).

Para que los procesos de una empresa generen productos de calidad es necesario obtener datos sobre distintas características tanto del proceso como del producto y analizarlos. La generación de dichos datos depende del tipo de característica que se considere, por ejemplo, si estamos hablando de diámetro, altura, peso etc. se deberá utilizar un instrumento de medición con el cual se generarán datos continuos. Ahora bien, si se considera como característica de interés un aspecto en el que sea complicado definir límites de especificación, y entonces se realice una inspección visual con un criterio del tipo pasa o no pasa, se obtendrán datos discontinuos. De acuerdo con Carro y González (2012) hay dos formas de evaluar la calidad: Una consiste en medir los atributos o las características del producto o servicio dónde es posible contar rápidamente para saber si la calidad es aceptable. (...). La otra forma de evaluar la calidad consiste en medir variables; es decir, las características del producto o servicio que son susceptibles de ser medidas, como peso, longitud, volumen o tiempo.

Para la representación y análisis de los datos, en busca de comprender y/o controlar el proceso de producción o la calidad del producto se han utilizado distintas herramientas en la industria. La Sociedad Americana de Calidad (ASQ) reconoce siete herramientas esenciales para control de calidad:

1. Diagrama de causa y efecto
2. Diagrama de Pareto
3. Hoja de verificación
4. Diagrama de dispersión
5. Histograma
6. Gráficos de control
7. Análisis de tendencias

Algunos de estas herramientas se describen en la sección 2.1.5.2.

Otra información importante que se puede obtener al analizar los datos generados de la medición de variables es su tendencia central y variabilidad.

Las *medidas de tendencia central* proporcionan información acerca del dato al cual se aproximan todos los del conjunto, las tres medidas principales son:

- **Media:** Es igual al promedio aritmético de un conjunto de datos, que se obtiene al sumarlos y el resultado se divide entre el número de datos (Gutiérrez y De la Vara, 2013).

Si se considera una muestra de tamaño n , $x_1 + x_2 + \dots + x_n$, entonces la media muestral \bar{X} está dada por:

$$\bar{X} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum_{j=1}^n x_j}{n} \quad (1)$$

Mientras que la media poblacional μ , se obtiene utilizando todos los datos de la población, más que una muestra como para la media muestral.

- **Moda:** es el dato que más se repite en el conjunto.
- **Mediana:** Es igual al valor que divide a la mitad a los datos cuando son ordenados de menor a mayor (Gutiérrez y De la Vara, 2013).

Por su parte para estudiar la variabilidad o dispersión se utilizan las siguientes medidas:

- Desviación estándar muestral, S : medida de la variabilidad que indica qué tan esparcidos están los datos con respecto a la media (Gutiérrez y De la Vara, 2013).

Dada una muestra de tamaño n , $x_1 + x_2 + \dots + x_n$, S está definida por:

$$S = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n - 1}} = \sqrt{\frac{\sum_{i=1}^n x_i^2 - \frac{(\sum_{i=1}^n x_i)^2}{n}}{n - 1}} \quad (2)$$

- Desviación estándar poblacional o del proceso, σ : se obtiene de una forma análoga a S , solo que utilizando todos los datos de la población y no solo una muestra (Gutiérrez y De la Vara, 2013).
- Rango: medición de la variación de un conjunto de datos que es resultado de la diferencia entre el dato mayor y el dato menor de tal conjunto (Gutiérrez y De la Vara, 2013).
- Coeficiente de variación: Medida de variabilidad que indica la magnitud relativa de la desviación estándar en comparación con la media (Gutiérrez y De la Vara, 2013).

2.1.3. Análisis de Componentes Principales

El análisis de componentes principales (Principal components analysis, PCA) es un método estadístico multivariante que consiste en reducir un conjunto grande de variables en uno de menor dimensión. Santos (2012) menciona que dicha técnica, se centra en la transformación ortogonal de un conjunto de datos correlacionados para obtener una combinación lineal de variables llamada componentes principales. Estos son los vectores propios de $XX = X^T X$, generalmente para la media centrada X y a veces después del auto escalamiento. PCA sigue la expresión,

$$X = T_A \cdot P_A^T + E_A \quad (3)$$

donde T_A es la matriz de puntuación $N \times A$, P_A es la matriz de carga $M \times A$ y E_A es la matriz de residuos $N \times M$. La puntuación de una nueva observación en el subespacio PCA se calcula de la siguiente manera:

$$t_n = x_n \cdot P_A \quad (4)$$

Donde x_n es un vector $1 \times M$ que representa una nueva observación y t_n es un vector

$1 \times A$ con las puntuaciones correspondientes, mientras que

$$e_n = x_n - T_n \cdot P_A^T \quad (5)$$

representa los residuales. Tanto los puntajes como los residuos se monitorean en un sistema MSPC. Si x es un vector de 1 a m características de calidad con valores propios: $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$ entonces la varianza de las componentes principales viene dada por su valor propio y la proporción de la varianza explicada, se determina como:

$$\frac{\lambda_j}{\lambda_1 + \lambda_2 + \dots + \lambda_m}$$

2.1.3.1. Metodología para seleccionar el número de componentes principales

Existen muchos métodos para decidir el número de componentes principales que se deben seleccionar. Rencher (2002) propuso los siguientes cuatro métodos:

Método 1 o Porcentaje: Esta técnica conserva las componentes suficientes que garanticen al menos un porcentaje específico de la proporción acumulada de la varianza total. Generalmente, el valor fijado es del 80%.

En este método, el reto está en seleccionar un porcentaje de umbral apropiado. Si se selecciona un valor de umbral muy alto, se corre el riesgo de incluir componentes que no generalicen a la población o variables específicas que no representan un resumen compuesto de varias variables.

Método 2 o Promedio: Este método consiste en conservar las componentes principales cuyos valores propios son mayores que el promedio de los valores propios, $\sum_{i=1}^p \frac{\lambda_i}{p}$. Por lo tanto, el método 2 elige aquellas componentes que explican más varianza que la varianza promedio de las variables. Para una matriz de correlación este promedio es 1.

Método 3 o Scree: El gráfico Scree es un procedimiento visual que representa el tamaño del valor propio en todo el número de valor propio. Es decir, se construye una gráfica de λ_i versus i y se busca una ruptura natural entre los valores grandes y los valores pequeños. Por lo tanto, la utilización de este enfoque permite determinar qué componentes son significativos aparte de la línea recta formada por los últimos valores propios. Desde el punto de vista práctico, puede suceder

que no exista de forma clara tal ruptura. Como resultado, este enfoque no es concluyente.

Método 4 o test de Bartlett (1954): Este método es una prueba estadística diseñada para ignorar los componentes principales que no son significativamente diferentes del resto y asume la normalidad multivariable. Por lo tanto, este método es una prueba de significancia de las componentes correspondientes a los valores propios más grandes. Generalmente, la cantidad de componentes principales producidas mediante este procedimiento son mayores a las que se obtienen por los métodos anteriores (Santos, 2012).

2.1.4. Índices de capacidad de procesos univariados y multivariados.

Los índices de capacidad de proceso (Process capability índices, PCI) son indicadores estadísticos muy utilizados en la industria para evaluar y describir la capacidad de un proceso. Los PCI cuantifican la capacidad de un proceso para alcanzar resultados dentro de un límite de especificación (SL) inferior y superior (de-Felipe & Benedito, 2017b). En general, esta cuantificación viene dada por un valor único (Chakraborty y Chatterjee, 2016). En este contexto, los PCI son indicadores de la bondad del proceso relacionado con la posición y la variabilidad de las medidas de las características del producto dentro de los SL. Un PCI con un valor "alto" indica que el proceso es capaz de producir un artículo que con toda probabilidad cumplirá o superará el requisito de los clientes (Chakraborty y Chatterjee, 2016). En el análisis de capacidades, se puede distinguir entre PCI univariados y multivariados. En un inicio, los PCI univariados se utilizaban con el objetivo de describir una sola característica del proceso. Considerando la mejora continua de los procesos de producción, la calidad del producto se analiza teniendo en cuenta varias características del producto de manera simultánea. Luego, los procesos de producción univariados se han convertido en procesos de producción multivariados. Por lo tanto, lo anterior ha provocado la introducción de PCI multivariados en la literatura. No obstante, la mayoría de estos no han sido aplicados de forma extensa en la industria (de-Felipe y Benedito, 2017b). Como consecuencia, en la mayoría de las empresas, los PCI univariados se utilizan

para para evaluar y describir la capacidad de los procesos multivariados al obtener un PCI univariado para cada característica del proceso (de-Felipe y Benedito, 2017b). De esta forma, al seguir este procedimiento es posible que se tenga como resultado una mala interpretación del proceso debido a que no se tiene en cuenta la influencia de la característica analizada sobre las otras características del producto en el proceso de producción multivariado analizado. Por tal motivo, cuando se analiza la capacidad de los procesos multivariados, los PCI univariados deben complementarse con los PCI multivariados ya que estos son capaces de describir de manera global la capacidad de un proceso multivariado teniendo en cuenta todas las características del producto simultáneamente (de-Felipe y Benedito, 2017b).

A continuación, se describen de forma breve algunos de los más importantes PCI univariados y multivariados existentes en la literatura.

2.1.4.1. Índices de capacidad de proceso univariado.

La capacidad del proceso se puede concebir como el campo del control de calidad centrado en la determinación de la viabilidad por el proceso para cumplir con las especificaciones (Santos, 2012). En particular, los PCI univariantes son indicadores estadísticos que se utilizan para cuantificar la bondad de un proceso al relacionar la variabilidad de las medidas de una característica de un solo producto con la admisible (de-Felipe y Benedito, 2017b). Generalmente, la capacidad del proceso se expresa en proporciones o índices entre las tolerancias y el rendimiento del proceso (Santos, 2012). Se dice que un proceso es capaz cuando casi todas las muestras se encuentran entre los SL. La mayoría de los estudios de capacidad consideran que las medidas de las características del producto obtenidas siguen una distribución normal (Montgomery, 2009). Por lo tanto, su ancho (variabilidad) se puede describir con la varianza (σ^2) de la distribución de características del producto. Los límites de tolerancia natural se colocan 3σ por encima y por debajo de la media(μ) de tal forma que aproximadamente el 99.73% de los valores extraídos de una distribución normal están dentro de este intervalo(6σ) (ver figura 2).

Con el objetivo de comprender las diferencias entre los PCI que serán presentados se introducen dos conceptos muy importantes: la región del proceso (PR) y la región de

especificación (SR). La PR se define como el intervalo que incluye el 99.73% de los valores extraídos de una distribución normal centrada en el valor medio de la característica del producto medido. Mientras tanto, la SR se define como el intervalo limitado por el límite inferior de especificación (LSL) y el límite superior de especificación (USL). Luego, la PR abarca el intervalo que se encuentra entre $\mu - 3\sigma$ y $\mu + 3\sigma$.

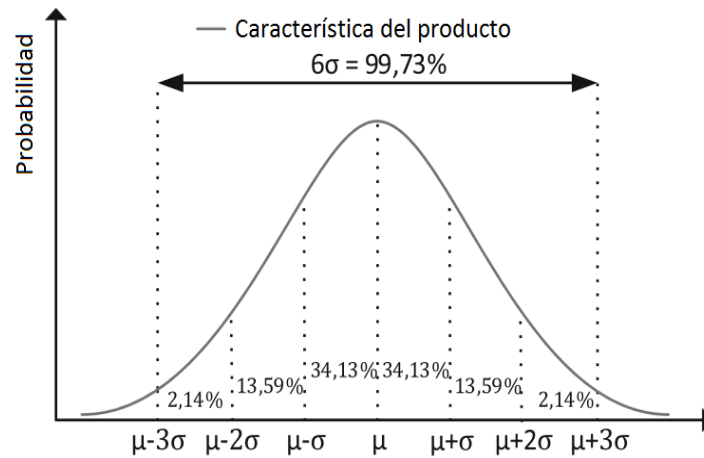


Figura 2. Anchura de las medidas distribuidas normales de una característica del producto (adaptada de (de- Felipe & Benedito, 2017)).

En la literatura se han propuesto muchos índices para medir la capacidad, siendo los más reconocidos los siguientes: C_p , C_{PU} , C_{PL} , k , C_{pk} , C_{Pm} y C_{Pmk} . De acuerdo con Gutiérrez y De la Vara (2013) son del tipo valor nominal es mejor, en donde, para considerar que hay calidad las mediciones deben ser iguales a cierto valor nominal o ideal (N), o al menos tienen que estar con holgura dentro de las especificaciones inferior (EI) y superior (ES).

El índice C_p

El índice C_p es un PCI univariado que muestra si el proceso analizado puede ser capaz. Este índice, compara el ancho de SR (USL - LSL) y el ancho de PR (6σ). La figura 3 muestra ambas regiones. El índice C_p se puede calcular mediante la siguiente expresión:

$$C_p = \frac{USL - LSL}{6\sigma} \quad (6)$$

Tenga en cuenta que la extensión de 6σ del proceso es la definición básica de la capacidad del proceso o la variación real. Sin embargo, σ es generalmente desconocido. Por lo tanto, debemos reemplazarlo con una estimación. Con frecuencia la desviación estándar ($s = \sqrt{\sigma^2}$) de una muestra de datos se utiliza como una estimación de σ . De acuerdo a los valores resultantes al calcular el índice C_p , es posible distinguir entre varios niveles de capacidad: muy excelente para valores C_p superiores a 2.00; excelente para valores entre $1.67 \leq C_p < 2.00$; satisfactorio para valores entre $1.33 \leq C_p < 1.67$; capaz para valores entre $1.00 \leq C_p < 1.33$; inadecuado para valores entre $0.67 \leq C_p < 1.00$; y pobre para valores de C_p menores a 0.67 (Tsai y Chen, 2006). En particular, si $C_p > 1$, la variabilidad real del proceso o la PR es menor que la PS del proceso. Como resultado, este proceso produce relativamente pocas unidades no conformes. Si $C_p = 1$, la variabilidad real del proceso es igual a la variabilidad tolerada representada por la PR. Para una distribución normal, esto implicaría alrededor de 0.27% (o 2700 ppm) de unidades no conformes. Finalmente, para un $C_p < 1$; el proceso presenta una variabilidad real mayor que la variabilidad tolerada. Por lo tanto, el proceso es muy sensible al rendimiento y se producirá un gran número de unidades no conformes.

El índice C_p asume que el proceso se centra en el punto medio de la región de especificación (figura 3). Sin embargo, en muchas situaciones lo anterior no se cumple y es posible tener un gran porcentaje de productos con características fuera de los SL, aunque el valor de C_p sea alto (de-Felipe & Benedito, 2017). Por lo tanto, fue necesario realizar alguna modificación de la PR que permita describir esta situación adecuadamente. Para ello, se introdujeron los índices C_{pU} , C_{pL} , k y C_{pk} respectivamente.

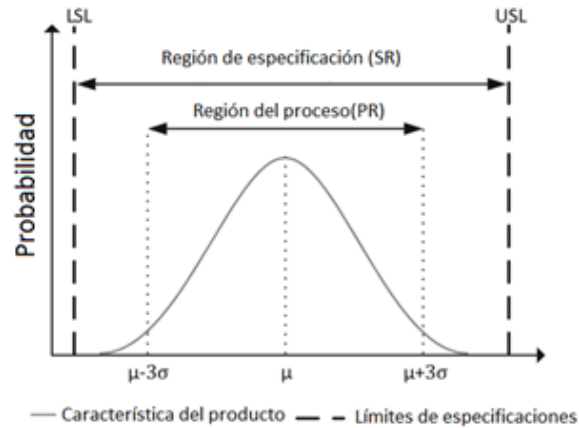


Figura 3. Regiones de proceso y especificación ((adaptada de (de-Felipe & Benedito, 2017)).

Los índices C_{PU} y C_{PL} .

El índice C_{PU} describe la relación entre la mitad superior de la SR ($USL - \mu$) y la mitad PR (3σ). El índice C_{PL} describe la relación entre la mitad inferior de la SR ($\mu - LSL$) y la mitad PR. Las expresiones de ambos índices se muestran a continuación:

$$C_{PU} = \frac{USL - \mu}{3\sigma} \quad (7)$$

$$C_{PL} = \frac{\mu - LSL}{3\sigma} \quad (8)$$

Las estimaciones \widehat{C}_{PU} y \widehat{C}_{PL} se obtendrían al reemplazar μ y σ en las ecuaciones (7) y (8) por estimaciones $\hat{\mu} = \bar{x}$ y, $\hat{\sigma} = s$ respectivamente.

Los índices C_{pk} y k

El índice C_{pk} considera la distancia mínima entre el punto medio de la PR y su más cercano SL. Este índice es el mínimo valor entre los índices C_{PU} y C_{PL} (ver ecuación 9).

$$C_{pk} = \min(C_{PU}, C_{PL}) \quad (9)$$

Por lo tanto, el índice C_{pk} relaciona la distancia escalada entre la media del proceso (o el valor medio de la característica del producto) y el límite de especificación más cercano. Generalmente, si $C_p = C_{pk}$, el proceso se centra en el punto medio de las especificaciones, y cuando $C_p > C_{pk}$ el proceso está descentrado. Por lo tanto, se concluye que la magnitud de C_{pk} en relación con C_p es una medida directa de cuán fuera de centro está operando el proceso.

El índice k mide la diferencia entre la media del proceso, μ , y el valor objetivo o nominal, N (target), para la correspondiente característica de calidad; y compara esta diferencia

con la mitad de la amplitud de las especificaciones (Gutiérrez y De la Vara, 2013). Dicho índice se calcula de la siguiente manera:

$$k = \frac{|T - \mu|}{\frac{USL - LSL}{2}} \quad (10)$$

El signo de valor absoluto se agregó para el caso cuando $LSL \leq \mu \leq T$.

Cuando el valor objetivo (T) es el punto medio de la SR, los índices C_{pk} y k se relacionan mediante la siguiente expresión:

$$C_{pk} = C_p(1 - k) \quad (11)$$

Si el valor medio de la característica del producto se encuentra exactamente en el centro de la SR, los índices C_{pk} y C_p tienen el mismo valor. Si el índice $C_{pk} > 1$, el proceso se define como capaz. En general, los índices C_{pk} y C_p son los PCI más conocidos para los casos univariados que son ampliamente adoptados por la industria (Hadian y Rahimifard, 2019).

El índice C_{pm}

Chang, Cheng, y Spiring (1988) introdujeron el índice C_{pm} (ver ecuación 12). Este índice también considera la posibilidad de que el valor objetivo no sea el punto medio de la SR. Por este motivo, los autores antes citados proponen una modificación de la PR. La nueva PR se encuentra en el intervalo $[T - 3\sigma', T + 3\sigma']$ donde $\sigma' =$

$\sqrt{\frac{\sum_{i=1}^n (x_i - T)^2}{n-1}}$. La nueva PR también incluye el 99.73% de los valores extraídos de una distribución normal, pero centrada en el valor objetivo. La varianza estimada σ'^2 se calcula considerando la distancia entre la característica del producto y el valor objetivo.

$$C_{pm} = \frac{\min(USL - T, T - LSL)}{3\sigma'} \quad (12)$$

En la figura 4a se muestra un ejemplo de la PR original y modificada, así como el SR. En este caso, $\mu + 3\sigma$ y $T + 3\sigma'$ están superpuestos.

El índice C_{pmk}

El índice C_{pmk} es una combinación de los índices C_{pk} y C_{pm} . Pearn, Kotz y Johnson (1992) introdujeron el índice C_{pmk} (ver ecuación 13). Este índice considera la posición del valor medio de la característica del producto dentro de la SR y también supone que el valor objetivo no está centrado en la SR. La nueva PR' se encuentra en el intervalo

$[\mu - 3\sigma', \mu + 3\sigma']$. Analizando las PR del índice C_{Pm} y del C_{Pmk} obtenemos como resultado que ambas PR tienen el mismo ancho. Sin embargo, la nueva PR' está centrada en el valor medio de las medidas como la PR del índice C_{Pk} (ver figura 4b). El valor de este nuevo índice está influenciado no solo por el ancho de la PR sino también por su posición. Como se puede observar en esta figura, la mitad superior del SR es más pequeña que la mitad superior de la PR modificada. Luego, el valor del índice C_{Pmk} es inferior a 1 y, por lo tanto, el proceso no es capaz. Sin embargo, los valores de los índices C_{Pm} y C_{Pk} son mayores a 1, por lo que, respecto a estos índices, el proceso es capaz (de-Felipe y Benedito, 2017b).

$$C_{Pmk} = \frac{\min(USL - \mu, \mu - LSL)}{3\sigma'} \quad (13)$$

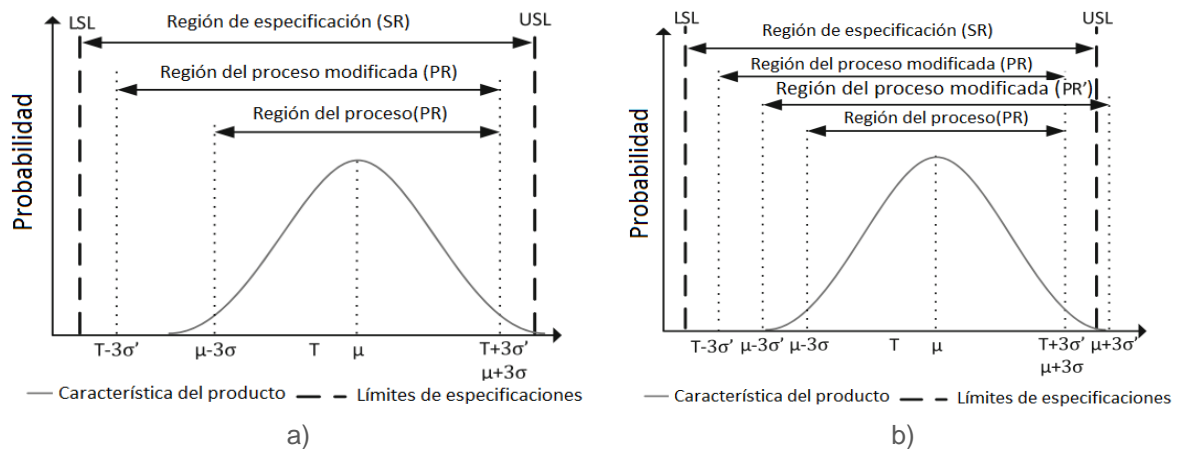


Figura 4. PR y SR modificada según K. Chang et al. (1988) (a), Modificación PR' según L. Pearn et al. (1992) (b).

2.1.4.2. Índices de capacidad de proceso multivariable (MPCI)

Los PCI multivariables son indicadores estadísticos que se utilizan para cuantificar la unión de un proceso multivariado con un solo índice para relacionar la variabilidad de las medidas de las características de múltiples productos con la admisible (de-Felipe y Benedito, 2017b).

Al igual que en el caso univariado, se supone que las medidas de las características del producto siguen una distribución normal. En el caso bivariado (y v -variado) a

diferencia del univariado, la PR se describe como la superficie (forma v -dimensional) que incluye el 99.73% de los valores extraídos de una distribución binormal (o v - multinormal) centrada en el valor medio de la muestra de datos. Matemáticamente, esta región se representa por lo siguiente expresión,

$$(X - \mu)^T \Sigma^{-1} (X - \mu) = c^2 \quad (14)$$

donde μ y Σ son el vector medio y la matriz de varianza y covarianza, c^2 depende de v .

Por lo general, se supone que la ecuación (14) sigue una distribución χ^2_v con v grados de libertad. Por lo tanto, alrededor del 99.73% de los valores extraídos de una distribución χ^2_v con v grados de libertad se encuentran dentro de la región delimitada por la ecuación (14). La región SR es la región paralelepípedica cuyos vértices están delimitados por los SL. La Figura 5 muestra la distribución de una muestra de datos para dos características del producto, la PR y la SR.

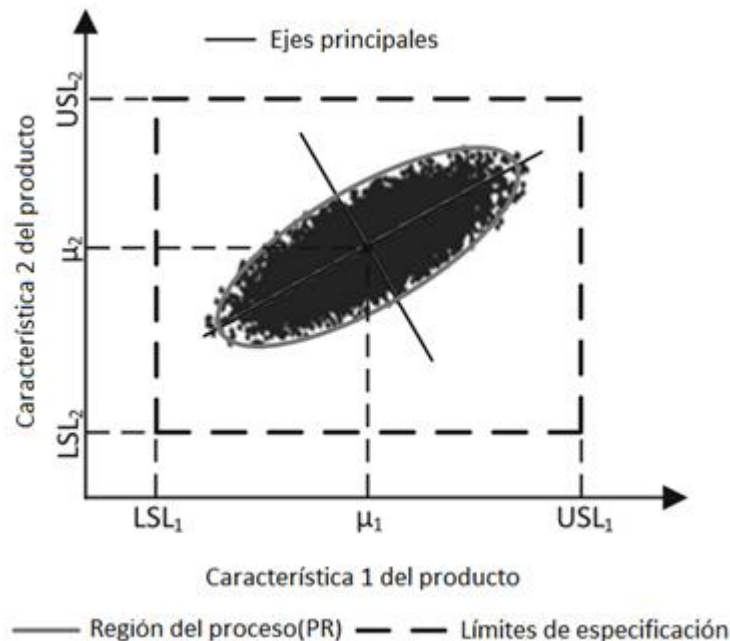


Figura 5. Representación de los ejes principales de la PR.
Pearn et al. (1992) (b).

Cuando se analiza la capacidad de los procesos de producción multivariados la correlación entre las características del producto juega un papel muy importante. En particular, la forma elipsoidal que representa la PR es más o menos inclinada,

dependiendo del valor de la correlación entre las características del producto analizado. Si no hay correlación entre las medidas, los ejes de la PR son paralelos a los ejes cartesianos. Por lo tanto, para evaluar adecuadamente la capacidad de los procesos de producción multivariados es necesario tomar en cuenta la correlación entre las características del producto. En la literatura los PCI multivariados se presentan y clasifican en dos grupos: los PCI multivariados que no tienen en cuenta la correlación entre las características del producto medido y los PCI multivariados que sí tienen en cuenta esta correlación. Dentro de estos últimos se destacan tres grupos:

- PCI multivariados basados en el análisis de componentes principales.
- PCI multivariados basados en la relación entre PR y SR.
- PCI multivariados basados en la función inversa de la función de distribución acumulada.

En este trabajo nos enfocaremos en los PCI multivariados basados en el análisis de componentes principales.

2.1.4.2.1. PCI multivariados basados en el análisis de componentes principales

Al igual que las PCI univariados, varios PCI multivariados también se han introducido recientemente en la literatura. Sin embargo, la mayoría de ellos aún no han sido muy aceptados por la industria (de-Felipe y Benedito, 2017a, 2017b). En los últimos años se han propuesto muchos índices basados en el análisis de componentes principales (PCA) como una extensión de los índices univariados C_P , C_{Pm} , C_{Pk} y C_{Pmk} . Algunos de los más aceptados son los índices sugeridos por F. Wang y Chen (1998), Xekalaki y Perakis (2002), y Wang (2005). El método PCA, permite obtener los ejes principales (vectores propios) de las PR (elipses, elipsoides, formas elipsoidales) y una matriz de varianzas y covarianzas diagonal que describe la PR eliminando la correlación entre características del producto medido. Como resultado, las características del producto modificadas derivadas de PCA no están correlacionadas. La Figura 6 muestra los vectores propios (u_i) que describen los ejes principales de la elipse en un caso bidimensional. Ambos ejes son ortogonales y no correlacionados.

F. Wang y Chen (1998) propusieron utilizar el análisis de componentes principales (PCA) basado en el teorema de Tong (Tong, 2012) para obtener PCI multivariados.

Por lo tanto, los PCI multivariantes presentados por estos autores (MC_P , MC_{Pm} , MC_{Pk} y MC_{Pmk}) se calculan encontrando los ejes de la PR y calculando los PCI univariados (C_P , C_{Pm} , C_{Pk} y C_{Pmk}) en la dirección de cada eje. Luego, se calcula la media geométrica de todos los PCI univariados y no correlacionados. Como este enfoque comienza con un PCA, se obtienen las variables no correlacionadas y se permite la reducción de la dimensionalidad.

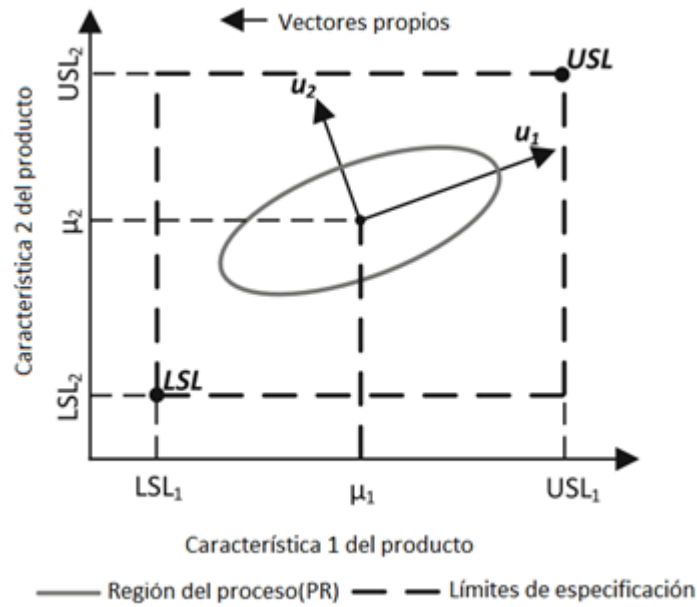


Figura 6. Descripción de PR con sus vectores propios. Pearn et al. (1992) (b).

Estos índices se basan en la descomposición espectral de la matriz de covarianza.

$$\Sigma = UDU^T \quad (15)$$

donde U es la matriz de vectores propios y D la matriz diagonal de los valores propios.

$$D = \text{Diag}(\lambda_1, \lambda_2, \dots, \lambda_p) \quad (16)$$

La i -ésima componente principal resulta $PC_i = u_i^T x$ y las especificaciones de ingeniería (SR y los valores objetivo) se transforman utilizando las siguientes ecuaciones,

$$LSL_{PC} = U \cdot LSL \quad (17)$$

$$USL_{PC} = U \cdot USL \quad (18)$$

$$T_{PC} = U \cdot T \quad (19)$$

donde, $LSL_{PC} = (LSL_{PC_1}, LSL_{PC_2}, \dots, LSL_{PC_v})$, $USL_{PC} = (USL_{PC_1}, USL_{PC_2}, \dots, USL_{PC_v})$ y $T_{PC} = (T_{PC_1}, T_{PC_2}, \dots, T_{PC_v})$ representan los vectores de los límites de especificación inferior y superior así como los valores objetivos correspondientes a cada componente principal. Mientras tanto, $LSL = (LSL_1, LSL_2, \dots, LSL_v)$, $USL = (USL_1, USL_2, \dots, USL_v)$ y $T = (T_1, T_2, \dots, T_v)$ representan los vectores de los límites de especificación inferior y superior así como los valores objetivos correspondientes a cada característica del producto. Finalmente, U es la matriz de rotación con los vectores propios de la matriz de varianza y covarianza.

La propuesta de K. Chang et al. (1988) viene dada por la siguiente expresión,

$$MC_p = \left(\prod_{i=1}^v C_p; PC_i \right)^{\frac{1}{v}} \quad (20)$$

donde,

$$C_p; PC_i = \frac{USL_{PC_i} - LSL_{PC_i}}{6\sigma_{PC_i}} \quad (21)$$

con v el número de componentes principales y $\sigma_{PC_i} = \sqrt{\lambda_i}$.

Igualmente MC_{Pk} , MC_{Pm} y MC_{Pmk} se obtienen reemplazando $C_p; PC_i$ por $C_{Pk}; PC_i$, $C_{Pm}; PC_i$ y $C_{Pmk}; PC_i$ respectivamente, donde:

$$C_{Pk}; PC_i = \min \left\{ \frac{USL_{PC_i} - \mu}{3\sigma_{PC_i}}, \frac{\mu - LSL_{PC_i}}{3\sigma_{PC_i}} \right\} \quad (22)$$

$$C_{Pm}; PC_i = \frac{USL_{PC_i} - LSL_{PC_i}}{6\sqrt{\sigma_{PC_i}^2 + (\mu - T)^2}} \quad (23)$$

y

$$C_{Pmk}; PC_i = \frac{\frac{USL_{PC_i} - LSL_{PC_i}}{2} - \left| \mu - \left[\frac{USL_{PC_i} + LSL_{PC_i}}{2} \right] \right|}{3\sqrt{\sigma_{PC_i}^2 + (\mu - T)^2}} \quad (24)$$

Wang y Du (2000) mejoraron el enfoque de Wang y Chen (1998) al disminuir el número de vectores propios necesarios que deben considerarse para describir la capacidad de un proceso multivariado. La contribución de Wang y Du (2000) radicó en seleccionar aquellos vectores propios que contribuyen a la mayor parte de la variabilidad del proceso considerando la relación de cada valor propio a la suma de los valores propios, que describe la proporción de variabilidad asociada con cada variable componente principal. Estos autores, proponen estimadores de intervalo de MC_p y MC_{pk} para un proceso multivariado con datos normales. Por otro lado, Wang (2005) sugiere otra forma de ponderar las componentes principales utilizando la media geométrica ponderada. Los índices propuestos dan lugar a:

$$MWC_p = \left(\prod_{i=1}^v C_p^{\lambda_i}; PC_i \right)^{\frac{1}{\sum_{i=1}^v \lambda_i}} \quad (25)$$

y así sucesivamente para MWC_{pk} , MWC_{pm} y MWC_{pmk} .

Chan, Cheng, y Spiring (1991) proponen una forma elipsoidal para representar la SR. Por su parte, Taam, Subbaiah, y Liddy (1993) modificaron la SR propuesta además de proponer el índice de capacidad de proceso multivariado MC_{pm} . Chen (1994) propone un método para estimar MC_p utilizando un enfoque de proporción no conforme. Shahriari, Hubele y Lawrence (1995) introdujeron un vector de capacidad para indicar la capacidad del proceso multivariado basado en el trabajo pionero de Hubele, Shahriari, y Cheng (1991). Pan y Lee (2010), analizan un caso especial del índice propuesto por Taam et al. (1993). Los autores señalaron que el índice propuesto por Taam et al. (1993) podría sufrir una sobreestimación si las características de calidad no son independientes. Más PCI multivariadas introducidas en la literatura y sus especificaciones están representadas en (de-Felipe y Benedito, 2017b).

2.1.5. Control estadístico de procesos

De acuerdo con Boardman (como se citó en Camacho et al., 2016) el control estadístico de procesos (CEP) es una metodología desarrollada por Walter Andrew Shewhart y respaldada por Williams Edwards Deming en el siglo pasado, con un tremendo impacto en la industria manufacturera de los Estados Unidos y Japón.

Existen varias definiciones para el CEP, según Deming es una colección de herramientas para la resolución de problemas, y aplica el análisis estadístico para medir, monitorear y controlar procesos por medio de gráficos o cartas de control (Herrera, Hernández , Figueroa y De La Ossa, 2018). Por su parte, Carro y González (2012) definen CEP como la aplicación de técnicas estadísticas para determinar si el resultado de un proceso concuerda con el diseño del producto o servicio correspondiente. La aplicación del CEP busca mejorar la calidad de los productos reduciendo la variabilidad y manteniendo estable el proceso. Además, permite detectar cambios introducidos mediante la identificación de causas especiales. En las siguientes dos secciones se definen conceptos y herramientas estadísticas utilizadas en el CEP.

2.1.5.1. Variabilidad y causas de variabilidad

Shewhart observó que no pueden producirse dos partes con las mismas especificaciones, lo cual se debe, entre otras cosas, a las diferencias que se dan en la materia prima, a las diferentes habilidades de los operadores y las condiciones en que se encuentra el equipo (Herrera et al., 2018). Por lo tanto, aun cuando el proceso se desarrolle bajo las mismas condiciones siempre habrá variación. De acuerdo con Gutiérrez y De la Vara (2013) la variabilidad se refiere a la diversidad de resultados de una variable o de un proceso. Estadísticamente es una medida de la dispersión de las observaciones en torno a la media (Carro y Gonzales, 2012). La variabilidad en los procesos ocurre debido a que en ellos intervienen factores sintetizados a través de las 6M: materiales, maquinaria, medición, mano de obra (gente), métodos y medio ambiente (Gutiérrez y De la Vara, 2013). De forma general la variación puede deberse a: causas comunes y causas especiales. La variación por causas comunes (o por azar) es aquella que permanece día a día, lote a lote; y es aportada de forma natural por las condiciones de las 6 M (Gutiérrez y De la Vara ,2013). Son fuentes de variación puramente aleatorias, no identificables e imposibles de evitar mientras se utilice el procedimiento actual (Carro y González, 2012). Mientras que la variación por causas especiales (o atribuibles) es generada por situaciones o circunstancias especiales que no están de manera permanente en el proceso (Gutiérrez y De la Vara ,2013). Las

causas especiales (...) a menudo pueden ser identificadas y eliminadas si se cuenta con los conocimientos y condiciones para ello (Gutiérrez y De la Vara ,2013). Por lo tanto, cuando un proceso trabaja sólo con causas comunes de variación se dice que está en control estadístico o es estable (Gutiérrez y De la Vara ,2013). Es importante saber distinguir entre estos dos tipos de causas cuando se presentan en un proceso, debido a que, de lo contrario, se podrían cometer dos tipos de errores, ver figura 7.

Errores que se pueden cometer en la conclusión del procesos luego de la aplicación de cartas de control			
		Estado real del proceso	
		Fuera de control: Existen causas especiales.	Bajo control: Existen solo causas comunes.
Conclusión del proceso	Fuera de control: Se detectan causas especiales.		Tipo I
	Bajo control: Se detectan solo causas comunes.	Tipo II	

Figura 7. Tipos de errores en la conclusión de un proceso.

Si se cuenta con los conocimientos adecuados se podrán generar las conclusiones correctas del estado del proceso sin cometer cualquiera de los dos tipos de errores.

2.1.5.2. Herramientas estadísticas utilizadas en el CEP

Las principales herramientas utilizadas para la resolución de problemas del CEP son: histograma o diagrama de tallo y hoja; hoja de verificación; gráfico de Pareto; diagrama de causa y efecto; diagrama de concentración de defectos; diagrama de dispersión y gráfico de control (Montgomery, 2009). Con la aplicación de estas herramientas se consigue monitorear el proceso y reducir la variabilidad. A continuación, se proporciona una descripción de las primeras seis herramientas. Montgomery (2009) sugiere que los gráficos de control se encuentran entre las herramientas de control de gestión más importantes, por lo cual su descripción se realiza en una sección por separado.

Histograma

El Histograma es una visualización gráfica del patrón de variación de un conjunto de datos (ASQ, 2017), por lo tanto, esta herramienta permite ver la tendencia central, la dispersión y la forma de la distribución (Gutiérrez y De la Vara, 2013).

Hoja de verificación

La hoja de verificación es un formato construido para coleccionar datos de forma que su registro sea sencillo, sistemático y que sea fácil analizarlos (Gutiérrez y De la Vara, 2013). Puede parecer la herramienta más sencilla, pero su elaboración debe delegarse a personas con experiencia en el área o situación que se evaluará.

Diagrama de Pareto

El diagrama de Pareto es simplemente una distribución de frecuencia (o histograma) de datos de atributos organizados por categoría (Montgomery, 2009), que permite identificar los defectos, errores o problemas de mayor prioridad en un proceso. La viabilidad y utilidad general del diagrama está respaldada por el llamado principio de Pareto, conocido como “Ley 80-20” o “Pocos vitales, muchos triviales”, en el cual se reconoce que pocos elementos (20%) generan la mayor parte del efecto (80%), y el resto de los elementos propician muy poco del efecto total (Gutiérrez y De la Vara, 2013).

Diagrama de causa y efecto

El diagrama de causa y efecto, también conocido como diagrama de Ishikawa, fue creado por el Dr. Kaoru Ishikawa. Una vez que un defecto, error o problema se ha identificado y aislado para un estudio adicional, debemos comenzar a analizar las posibles causas de este efecto no deseado (Montgomery, 2009) para lo cual resulta útil esta herramienta. El diagrama de causa y efecto es un método gráfico que relaciona un problema o defecto con los factores o causas que posiblemente lo generan (Gutiérrez y De la Vara, 2013). Los factores de causa se organizan en categorías y se muestran en un diagrama (ASQ, 2017).

Diagrama de concentración de defectos

De acuerdo con Montgomery (2009) un diagrama de concentración de defectos es una imagen de la unidad, que muestra todas las vistas relevantes. Luego, los diversos tipos de defectos se dibujan en la imagen y el diagrama se analiza para determinar si la ubicación de los defectos en la unidad transmite información útil sobre las posibles causas de los defectos.

Diagrama de dispersión

El diagrama de dispersión es la representación gráfica en el plano X - Y, de los valores que toman dos variables. Si se observa que los puntos siguen algún patrón definido, esto será evidencia de una posible relación entre las dos variables (Gutiérrez y De la Vara, 2013). En consecuencia, con el gráfico podemos identificar patrones, tendencias o clúster en los datos .

2.1.5.3. Gráficos de Control

De las siete herramientas, el gráfico de control de Shewhart es probablemente el más sofisticado, fue desarrollado en 1920 por Walter A. Shewhart (Montgomery, 2009) . El gráfico de control es una visualización gráfica de una característica de calidad que se ha medido o computado a partir de una muestra en comparación con el número o el tiempo de la muestra (Montgomery, 2009) Los gráficos o cartas de control se especializan en estudiar la variabilidad a través del tiempo (Gutiérrez y De la Vara, 2013), así, permiten observar los resultados obtenidos periódicamente de una forma didáctica, indicando si las variaciones que se presentan en un proceso están por fuera de lo normal (Argumedo, Molina , Martínez y Hernández, 2017). Con la aplicación de los gráficos de control se busca identificar y eliminar causas especiales; reducir la variación debida a causas comunes, e identificar oportunidades de mejora en el proceso. Otra utilidad que tienen los gráficos de control es la de estimar los parámetros de un proceso de producción y, a través de esta información, determinar la capacidad del proceso (Montgomery, 2009).

Montgomery afirma que los elementos que componen un gráfico de control son:

- “Una línea central (CL) que representa el valor promedio de la característica de calidad correspondiente al estado en control. (Es decir, solo están presentes las causas casuales).
- Otras dos líneas horizontales, llamadas límite superior de control (UCL) y límite inferior de control (LCL), (...). Estos límites de control se eligen de modo que, si el proceso está en control, casi todos los puntos de muestra se encontrarán entre ellos” (2009, p. 182).

La determinación de los límites de control se debe realizar adecuadamente. Debido a que se podría incurrir en resultados equivocados, por ejemplo: considerar puntos

dentro de los límites de control cuando en realidad deben estar fuera, y con esto descartar la presencia de causas especiales, o viceversa, lo que nos llevaría a cometer un error del tipo I o del tipo II.

Consideremos W como la variable del proceso que nos interesa analizar, de acuerdo con Gutiérrez y De la Vara (2013), para el caso que W se distribuye normal con media μ_W y desviación estándar σ_W , y bajo condiciones de control estadístico se tiene que entre $\mu_W - 3$ y $\mu_W + 3$ se encuentran 99.73% de los posibles valores de W (figura 2). Por lo tanto, bajo las condiciones anteriores, el límite de control inferior (LCL), la línea central (CL) y el límite de control superior (UCL) (ver figura 6), quedan determinados como:

$$LCI = \mu_W - 3 \tag{26}$$

$$LC = \mu_W \tag{27}$$

$$LCS = \mu_W + 3\sigma_W \tag{28}$$

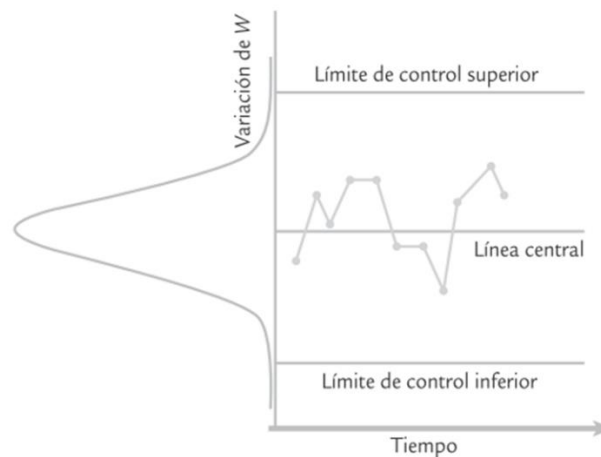


Figura 8. Componentes de un gráfico de control. Gutiérrez y De la Vara (2013).

Los gráficos de control se clasifican de forma general en dos tipos:

- Los gráficos de control de variables (tipo Shewhart): se utilizan cuando las variables o características de calidad son de tipo continuo, permiten monitorear la media y la variabilidad de los datos medidos. Entre los gráficos de control para variables se encuentran los siguientes:
 1. Gráficos de control de la media y el rango, $\bar{X} - R$.

2. Gráficos de control de la media y la desviación típica, $\bar{X} - S$.
 3. Gráficos de control para individuales
- Los gráficos de control por atributos: se aplican a variables o características de calidad del tipo pasa o no pasa
1. Gráficos de control para proporción de defectuosos, p .
 2. Gráficos de control para número de defectuosos, np .
 3. Gráficos de control para número de defectos, c .
 4. Gráficos de control de defectos por unidad, u .
- Otros gráficos de control utilizados son las que permiten la detección de cambios pequeños:
1. Gráficos de control de promedios móviles exponenciales ponderados (EWMA).
 2. Gráfico de control de sumas acumuladas (CUSUM).

Un último aspecto importante para considerar en la aplicación de los gráficos de control es el análisis de patrones seguidos por los puntos graficados. Debido a que la presencia de un patrón no aleatorio también indica una condición fuera de control. En la tabla 2, se muestran los principales tipos de patrones no aleatorios que se podrían encontrar en los gráficos de control.

Patrones no aleatorios que se pueden presentar en caratas de control

1	Un punto fuera de los límites de control
2	Nueve puntos consecutivos en el mismo lado
3	Seis puntos consecutivos ascendentes o descendentes
4	Catorce puntos consecutivos alternando arriba y abajo
5	Dos de tres puntos consecutivos situados más allá de dos sigmas.
6	Cuatro de cinco puntos consecutivos situados más allá de una sigma.
7	Quince puntos consecutivos situados a menos de un sigma
8	Ocho puntos consecutivos situados a más de un sigma (mismo lado)

Tabla 2. Patrones no aleatorios que se pueden encontrar en un gráfico de control.

Fase I y Fase II de la aplicación de gráficos de control

La aplicación de los gráficos de control se realiza en dos etapas: Fase I y Fase II.

La Fase I es también conocida como Fase de estabilización, debido a que estos estudios se utilizan cuando los cuadros de control se establecen por primera vez y con el objetivo de llevar el proceso al control estadístico (Santos, 2012). En esta fase se recopila un conjunto de datos del proceso y se analizan todos a la vez en un análisis retrospectivo, construyendo límites de control de prueba para determinar si el proceso ha estado en control durante el período de tiempo en que se recopilaron los datos, y para ver si son confiables. Se pueden establecer límites de control para monitorear la producción futura (Montgomery, 2009).

La Fase II o fase de monitoreo, comienza después de que tengamos un conjunto "limpio" de datos de proceso recopilados en condiciones estables y representativos del rendimiento del proceso en control (Montgomery, 2009). En esta fase, los cuadros de control se emplean para verificar si el proceso permanece en control. Aquí, la variabilidad del proceso se monitorea utilizando la media y la covarianza lograda desde la Fase I (Santos, 2012).

Para obtener un conjunto "limpio", Montgomery (2009) afirma que en la Fase I los puntos que están fuera de los límites de control se investigan, buscando posibles causas asignables. Todas las causas asignables que se identifican son trabajadas por el personal de ingeniería y operativo en un esfuerzo por eliminarlas. Los puntos que están fuera de los límites de control se excluyen y se calcula un nuevo conjunto de límites de control revisados. Luego se recopilan nuevos datos y se comparan con estos límites revisados(...). Finalmente, el proceso se estabiliza y se obtiene un conjunto limpio de datos que representan el rendimiento del proceso en control para su uso en la fase II.

La aplicación de las cartas de control para el control estadístico univariado de procesos nos permite analizar el estado del proceso respecto a una sola variable. Sin embargo, cuando se deben medir dos o más variables puede resultar ineficaz. Para dichos casos

resulta de mayor utilidad el MSPC. En la siguiente sección se abordan algunos conceptos y herramientas importantes del MSPC.

2.1.6. Control estadístico multivariado de procesos

Es común que en los procesos de producción se deban estudiar más de una característica de calidad, para lo cual, utilizar cartas de control univariantes para cada una de las variables definidas, de forma individual, generará una gran cantidad de cartas de control. Además, cuando suceden eventos especiales (difíciles de detectar mediante SPC univariado) en un proceso, ocurren afectaciones tanto en la magnitud de las variables como en la relación entre ellas. Por lo tanto, el monitoreo de dos o más características de calidad de forma independiente en un enfoque univariado puede conducir a resultados muy engañoso (Bersimis, Psarakis, y Panaretos, 2007). Como consecuencia, se requieren cartas de control multivariantes que tomen en cuenta de forma simultánea múltiples variables (Das y Prakash, 2008; Kim, Sukchotrat, y Park, 2011). El procedimiento de monitorear procesos en los que múltiples características relacionadas son de interés se conoce como SPC multivariado (MSPC), introducido por primera vez por Hotelling (1947). Por lo tanto, el MSPC puede ser visto como una extensión del SPC para controlar varias variables al mismo tiempo (Camacho et al., 2016).

En la literatura se han introducido diferentes cartas de control multivariadas. Entre las cuales se destacan: Las cartas de control T^2 de Hotelling, las cartas de suma acumulativa multivariable (MCUSUM) y las cartas de control de media móvil ponderada exponencial multivariable (MEWMA) (Lowry y Montgomery, 1995). De las cartas anteriores, la más utilizada es la primera carta (Mason y Young, 2002; Rogalewicz y Poznańska, 2013), la cual se puede considerar como una versión multivariable de la carta de Shewhart (Phaladuganon, Kim, Chen y Jiang, 2013). La carta T^2 de Hotelling utiliza una estadística T^2 que cuantifica la distancia entre los valores de las variables del proceso/producto y el vector de sus medias o el centro de observaciones en control (IC) conocidas o estimadas a partir de la muestra base (Park y Kim, 2018). Además, tiene en cuenta una estructura de correlaciones entre variables en forma de matriz de covarianza. Como mencionan Testik y Runger, el límite de decisión de un gráfico de

control T^2 tiene una forma elipsoidal que puede detectar un cambio de la media del proceso en el espacio multivariado en cualquier dirección (Park y Kim, 2018). La figura 9 muestra la diferencia entre una carta de control bidimensional y dos cartas de control univariadas.

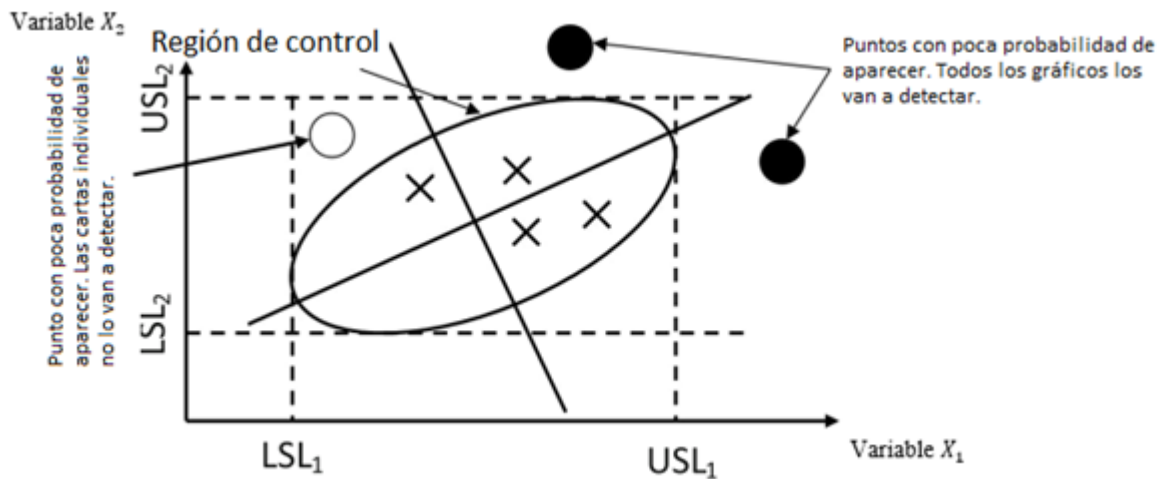


Figura 9. Cartas de control bidimensional en comparación con dos cartas de control univariadas (Adaptada de Mason y Young (2002)).

En la figura 9 se destaca la región de control o la PR (elipse). Esta es una forma análoga para controlar los límites en las cartas de control de Shewhart tradicionales. Como se observa esta región es más pequeña que el rectángulo formado por las dos cartas de control individuales ya que consideró una correlación entre las variables. Además, se puede ver que las cartas de control de Shewhart tradicionales no son capaces de detectar que el punto en blanco se encuentra visiblemente en contraposición con la correlación positiva en la que se basa la región de control. Si la variable X_1 toma un valor pequeño, la variable X_2 también debe de seguir el mismo comportamiento. Cuando lo anterior no sucede, se dice que estamos en presencia de un proceso que se salió de control. Luego de observar la figura 6, se puede concluir que las cartas de control multivariadas son más efectivas que las tradicionales para detectar algunos síntomas del proceso que nos indican que este está fuera de control (Rogalewicz y Poznańska, 2012). Agregar información de muchas variables en un gráfico constituye una de las grandes ventajas que presentan las cartas de Hotelling. En particular, el operador puede monitorear el proceso utilizando sólo una carta en vez

de muchas. Como resultado, puede ser capaz de asegurar con mayor facilidad la calidad del proceso. Sin embargo, la vigilancia multivariada también puede ser problemático ya que fuera de la región de control se pueden presentar algunas situaciones tales como: el proceso está fuera de control debido a una variable específica, debido a que dos o más variables actúan de forma simultánea o debido al cambio en la matriz de covarianza. Luego, el procedimiento que se sigue con el fin de establecer o conocer las causas de lo que realmente ocurrió con el proceso puede ser muy complicado. Por lo tanto, la principal desventaja en estas cartas se relaciona con la interpretación de los patrones que aparecen. Otra desventaja que encontramos en las cartas de Hotelling es que puede resultar difícil de entender sin una base matemática sólida ya que utiliza muchas herramientas de álgebra matricial. Finalmente, la T^2 de Hotelling no está en las unidades de ninguna variable particular, lo que puede ser un poco confuso la interpretación de sus resultados. Considerando estas desventajas, y teniendo en cuenta la facilidad, la implementación e interpretación de las cartas de control univariadas hacen que en la industria se apliquen con mayor popularidad que las cartas multivariadas. Sin embargo, puede suceder que el limitado uso en la industria de las cartas multivariadas se deba a la falta de una metodología clara que permita implementar las cartas de control multivariadas y los índices de capacidad (Rogalewicz y Poznańska, 2013).

En la actualidad, se hace más potente la implementación y el desarrollo de métodos computacionales y el uso de computadoras en el entorno de producción. Por lo tanto, el control de proceso estadístico multivariado no debería enfrentar tales barreras como hace algunos años. Considerando el hecho de que en muchos procesos es frecuente recopilar una gran cantidad de datos de forma automática y que generalmente estos datos no tienen ningún propósito debido a su complejidad y el escaso conocimiento que existe sobre que métodos utilizar para su análisis, se asume que existe una necesidad desarrollar herramientas SPC multivariantes y ajustarlas a la práctica industrial.

2.1.6.1. **Cartas de control multivariado para observaciones individuales y medias.**

Todos los procedimientos de MSPC asumen que no hay cambios en la matriz de covarianza. Por tal motivo, es muy importante recopilar una muestra base de tamaño suficiente y usar un estimador consistente de la matriz de covarianza. Teniendo en cuenta, si la muestra recolectada del proceso contiene una sola observación o algunas observaciones, es posible distinguir entre cartas de control de Hotelling para observaciones individuales o cartas para medias. A continuación, se presentan de forma breve algunas fórmulas relativas a estas dos cartas.

2.1.6.1.1. *Cartas de control multivariadas para medias.*

Se utilizan cartas de control multivariadas para las medias cuando existe la posibilidad de recolectar algunas muestras homogéneas de tamaño n que describen el proceso en el momento de la recolección (Rogalewicz y Poznańska, 2012). Frecuentemente, en la práctica, los parámetros μ y Σ son desconocidos y, por lo tanto, deben estimarse a través de estimadores no sesgados \bar{x} y S . Con base a la generalización multivariable del estadístico t a partir de la teoría normal univariada:

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}} \quad (29)$$

Haciendo

$$t^2 = \frac{(\bar{x} - \mu)^2}{S^2/n} = n(\bar{x} - \mu) (S^2)^{-1}(\bar{x} - \mu) \quad (30)$$

La generalización es:

$$T^2 = n(\bar{X} - \bar{\bar{X}})^t (S)^{-1}(\bar{X} - \bar{\bar{X}}) \quad (31)$$

Donde \bar{X} y S representan el vector de medias y la matriz de covarianzas, respectivamente. Las estadísticas T^2 siguen una distribución F con p y $(mn - m - p + 1)$ grados de libertad. Luego, los límites de control establecidos para las Fases I y II son los siguientes:

Límites de control para la fase I

Límites de control para la fase II

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

$$LCL = 0 \quad LCL = 0$$

De acuerdo con Lowry y Montgomery (1995), la aplicación de esta carta requiere varias características de calidad entre 2 y 10, y registros de más de 20 muestras (a menudo más de 50) de tamaño 2, 3 o 10.

2.1.6.1.2. Cartas de control multivariado para observaciones individuales.

En muchos procesos, debido a su propia naturaleza (no hay posibilidad o no hay explicación económica), solo es posible tomar una observación en cada intervalo de tiempo. Por ejemplo, cuando la tasa de producción es muy lenta, las mediciones son muy caras o son destructivas. Este caso es frecuentemente llamado para individuales. Significa que en la estructura de datos del proceso solo se registra una observación por variable en el tiempo m , por lo tanto, $n = 1$ (Santos, 2012). Aquí, la suposición común y más importante se relaciona con la distribución de datos, la cual debe corresponder a una normal multivariable (Rogalewicz y Poznańska, 2012).

Siendo $n = 1$, la estadística T^2 viene dada por, $T^2 = n(X - \bar{X})^t (S)^{-1} (X - \bar{X}) = (X - \bar{X})^t (S)^{-1} (X - \bar{X})$. Tracy, Young, y Mason (1992) definen los límites de control como:

Límites de control para la fase I

$$UCL = \frac{(m-1)^2}{m} \beta_{\alpha,p/2,(m-p-1)/2}$$

$$LCL = 0$$

Límites de control para la fase II

$$UCL = \frac{p(m+1)(m-1)}{m^2 - mp} F_{\alpha,p,m-m-p}$$

$$LCL = 0$$

Donde β es la distribución beta con $p/2$ y $(m-p-1)/2$ grados de libertad con un nivel de significación α .

2.1.6.1.3. Carta de control basada en el análisis de componentes principales (PCA)

Cuando el número de variables es muy grande y/o las variables están altamente interrelacionadas, se recomienda el uso de métodos de variables latentes, como PCA, en el MSPC (Camacho et al., 2016). En MSPC basado en PCA, es común el monitoreo de la estadística D o la estadística T^2 de Hotelling (Hotelling, 1947) que se calcula a partir de las puntuaciones. Después de calcular la estadística a partir de los datos de calibración, se puede establecer en la carta, los límites de control a un cierto nivel de confianza (MacGregor y Kourti, 1995; Tracy et al., 1992). Luego, los nuevos datos son monitoreados usando estos límites. Por lo tanto, cuando los límites se superan de manera significativa o consistente se dice que el proceso se encuentra fuera de control (Westerhuis, Gurden, y Smilde, 2000). De acuerdo con Camacho et al. (2016) la estadística D para la observación n se puede calcular a partir de la siguiente ecuación,

$$D_n = \sum_{a=1}^A \left(\frac{t_{an} - \mu_{t_a}}{\sigma_{t_a}} \right)^2 = \sum_{a=1}^A \frac{t_a^2}{\lambda_a} \quad (32)$$

donde t_{an} representa la puntuación de la observación en la a -ésima componente, μ_{t_a} y σ_{t_a} representan la media y la desviación estándar de las puntuaciones de esa componente en los datos calibrados, respectivamente. La estadística D o la T^2 de Hotelling representa la distancia de Mahalanobis (Praus, 2018) estimada desde el centro del subespacio latente hasta la proyección de una observación en este subespacio. Además, las puntuaciones son combinaciones lineales de las variables originales y, por lo tanto, de acuerdo con el teorema del límite central, se supone que tienen una distribución aproximadamente normal. Por lo tanto, al multiplicar la estadística D en la fase I por una constante obtenemos que esta sigue una distribución beta, β (Tracy et al., 1992). Luego,

$$D \sim \frac{(N-1)^2}{N} B_{\frac{A}{2}, (N-A-1)/2} \quad (33)$$

Por lo tanto, el límite superior de control (UCL) correspondiente para la estadística D con un nivel de significancia α viene dada por:

$$UCL(D)_\alpha = \frac{(N-1)^2}{N} B_{\left(\frac{A}{2}, (N-A-1)/2\right), \alpha} \quad (34)$$

Mientras que en la fase II, para los nuevos datos de entrada, la estadística D multiplicada por una constante sigue una distribución F (Tracy et al., 1992).

$$D \sim \frac{A(N^2-1)}{N(N-A)} F_{A, (N-A)} \quad (35)$$

Y UCL correspondiente a nivel de significancia α viene dada por:

$$UCL(D)_\alpha = \frac{A(N^2-1)}{N(N-A)} F_{A, (N-A), \alpha} \quad (31)$$

2.2. Antecedentes

En la mayoría de las empresas sigue siendo muy común utilizar el SPC. Sin embargo, ya se han desarrollado varias investigaciones en las que se muestra la implementación del MSPC basado en ACP. El campo de aplicación del MSPC basado en ACP ha sido variado, implementándose en áreas de computación, sociales, ecología y la industria, entre otras. Algunos de los trabajos más referenciados o realizados a lo largo del tiempo se mencionan a continuación.

- *Diagnosing network-wide traffic anomalies*. Autores: Lakhina A, Crovella M, Diot C. 2004. En su área, es quizás el trabajo más referido (Camacho et al., 2016).
- *The use of principal components and univariate charts to Control multivariate processes*. Autores: Marcela A. G. Machado, Antonio F. B. Costa. 2007. En este trabajo se realiza una comparación entre los gráficos T2 basado en componentes y los gráficos de control univariados simultáneos para cada una de las variables originales o los basados en las componentes principales. Concluyendo con que los gráficos de control univariados basados en las componentes principales son los que tienen mejor desempeño cuando existe una alta correlación positiva entre las variables.
- *Multivariate Statistical Process Control Based on Principal Component Analysis (MSPC-PCA): Some Reflections and a Case Study in an Autobody Assembly Process*. Autor: Alberto Ferrer. 2008. Muestra los beneficios de usar MSPC basado en PCA en los procesos de fabricación.

- Multivariate statistical process control in product quality review assessment- A case study. Autores: M. Kharbach, Y. Cherrah, Y. Vander Heyden, A. Bouklouze. 2017. Se evaluó un producto farmacéutico estudiando variables referentes a seis ingredientes activos. El estudio se realizó utilizando SPC Y MSPC basado en PCA. Con la aplicación de SPC resultó que todos los lotes estaban bajo control, mientras que MSPC arrojó que había de 4 a seis lotes fuera de control. Además, se obtuvo que con MSPC se pueden notar cambios sutiles en el proceso de producción del producto farmacéutico.
- *Principal components in multivariate control charts applied to data instrumentation of dams.* Autores: Emerson Lazzarotto, Liliana Madalena Gramani, Anselmo Chaves Neto, Luis Albino Teixeira Junior. 2016. Este trabajo realiza un análisis estadístico multivariante y propone un modelo que utiliza los componentes principales, las tablas de control y la estadística T^2 . Buscan establecer límites de control para las lecturas generadas por los instrumentos, con la intención de detectar valores atípicos, así como identificar los instrumentos que contribuyen a una mayor variabilidad.

2.3. El Control de procesos y el cumplimiento de los requisitos de las Normas.

En el sector automotriz la norma IATF 16949:2016 Sistemas de Gestión de la Calidad, su sección 9.1.1.1. que trata sobre el Seguimiento y medición de los procesos de fabricación, establece que:

“La organización debe #219 realizar estudios de proceso de todos los procesos de fabricación nuevos (incluidos los de ensamble o secuencia) para verificar la habilidad del proceso y proporcionar entradas adicionales para el control del proceso, incluidos aquellos para las características especiales.

La organización debe #220 mantener la habilidad o el desempeño del proceso de fabricación como está especificado en los requisitos del proceso de aprobación de piezas del cliente.(...)

Para las características que estadísticamente pierden su habilidad o son inestables, la organización debe #223 iniciar un plan de reacción a partir del

plan de control y evaluar su impacto en el cumplimiento de las especificaciones.
(...)

La organización debe #225 desarrollar e implementar un plan de acción correctiva, indicando las acciones específicas, los plazos y las responsabilidades asignadas, para asegurar que el proceso de nuevo es establece y estadísticamente hábil. (...).”

En la misma norma se establece en el numeral 9.1.1.2. que la empresa debe de establecer las técnicas estadísticas que utilizará e incluirlas en el Análisis de riesgos del diseño.

Para otros sectores de la industria, la norma ISO 9001-2015, en su numeral 9.1.2. establece que:

La organización debe #109 determinar:

- a) qué necesita seguimiento y medición;
- b) los métodos de seguimiento, medición, análisis y evaluación necesarios para asegurar resultados válidos;
- c) cuándo se deben #110 llevar a cabo el seguimiento y la medición;
- d) cuándo se deben #111 analizar y evaluar los resultados del seguimiento y la medición.

La organización debe #112 evaluar el desempeño y la eficacia del sistema de gestión de la calidad.

La organización debe #113 conservar la información documentada #28 apropiada como evidencia de los resultados.”

Por lo tanto, la aplicación de técnicas estadísticas para asegurar el control de los procesos queda bien establecido en el contenido de las normas de Sistemas de Gestión de Calidad.

Capítulo 3. Metodología

3.1. Materiales y métodos

Los materiales utilizados para desarrollar la investigación fueron:

- Documentación en línea, revistas y libros.
- Computadora con software R instalado.

Los métodos que se utilizaron para lograr cada uno de los objetivos específicos y finalmente el objetivo general de la investigación, se muestran a modo de resumen en la figura 10. Dichos métodos se aplicaron para la Fase I y II.

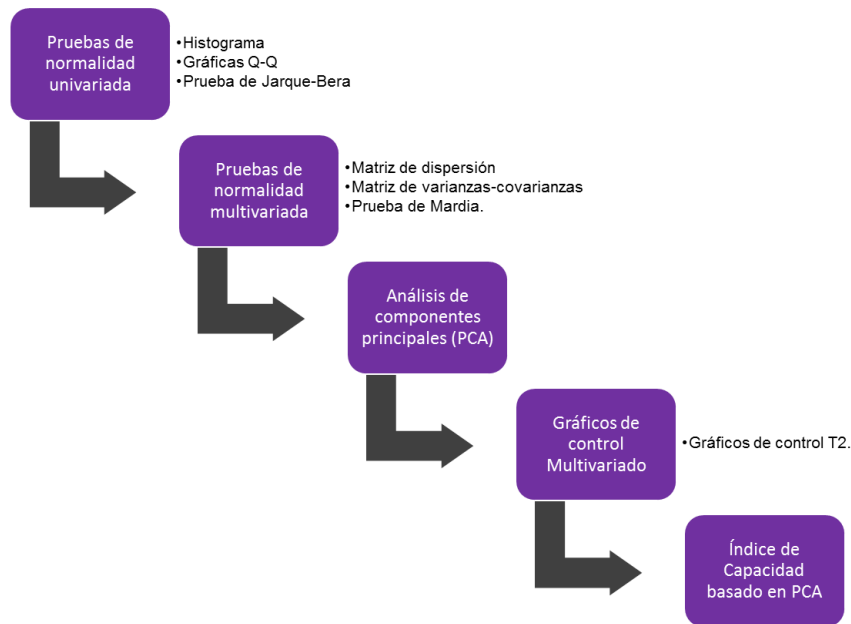


Figura 10. Métodos utilizados en el desarrollo la investigación.

3.2. Descripción del caso

En esta sección, discutiremos brevemente una aplicación de las técnicas de MSPC en la industria. Específicamente, analizaremos un caso real de cinco variables en relación con la calidad de un proceso en la fabricación de bimetales para un termostato. Las 5 características de calidad medidas fueron: deflexión, curvatura, resistividad y la dureza en los lados de expansión baja y alta (durezab, durezaa), de termostatos bimetálicos

de latón y acero. Los datos se dividen en dos grupos, correspondientes a las Fases I y II para la aplicación del MSPC. Ambos grupos de datos cuentan con 28 mediciones por cada una de las variables y se encuentran disponibles en los paquetes Bimetal1 y Bimetal2 del Software R. Los datos fueron organizados y tabulados para la aplicación del control estadístico multivariado. Por lo tanto, nuestra matriz de datos para cada fase es 28 x 5, donde el nivel de deflexión esta medido en $10^{-6}K^{-1}$, el nivel de curvatura esta medido en $10^{-6}K^{-1}$, el nivel de resistividad en $10^{-1}ohm \cdot \frac{mm^2}{m}$, la dureza del lado de baja expansión en $10 \frac{N}{mm^2}$ y la dureza del lado de expansión alta en $10 \frac{N}{mm^3}$. En las tablas 1 y 2 se muestran los resultados de las mediciones por cada variable con el fin de tener los datos para analizar la capacidad multivariada del proceso en la Fase I y en la Fase II respectivamente.

Variables medidas Fase I					
	Deflexión	Curvatura	Resistividad	Durezab	Durezaa
1	20.84	39.84	14.98	21.88	25.87
2	20.89	39.94	14.91	22.03	25.97
3	21.13	40.12	15.58	22.13	26.11
4	20.42	39.78	14.73	21.46	25.74
5	21.29	40.31	15.56	22.65	26.22
6	21.08	39.98	15.19	22.22	25.91
7	21.31	40.23	15.37	22.16	26.28
8	21.77	40.12	15.57	22.2	26.16
9	20.82	39.93	15.71	22.09	25.95
10	21.31	39.95	15.06	22.12	25.93
11	21.22	40.03	15.39	21.79	26.04
12	20.5	39.86	15.11	22.19	25.8
13	21.07	40.22	15.48	21.93	26.21
14	21.28	40.11	15.12	22.3	26.09
15	21.46	40	15.29	21.87	26.06
16	21.09	40.04	14.56	22.14	25.88
17	21.01	39.95	15.1	21.93	25.9
18	21.07	40.11	15.48	21.98	26.06
19	20.62	40.08	15.26	21.68	26.16
20	21.1	40.08	14.89	22	26.3
21	21.22	40.11	15.38	22.06	26.13
22	20.74	40.05	15.65	22.09	25.94
23	20.76	39.98	14.98	22.29	26.03
24	21	40.11	15.17	22.04	25.99
25	20.57	39.73	14.35	22.02	25.8
26	20.78	39.83	15.27	21.6	25.89
27	20.96	40.03	15.26	21.98	25.94
28	21.14	39.93	14.98	21.84	25.98

Tabla 3. Mediciones por variable para la Fase I.

Variables medidas Fase II					
	Deflexión	Curvatura	Resistividad	Durezab	Durezaa
1	21.07	39.98	15.11	22.12	25.95
2	21.38	39.89	15.19	21.85	25.82
3	21.4	40.06	15.82	21.86	26.16
4	20.49	39.87	14.56	21.95	26.06
5	21.08	39.94	14.73	21.98	25.93
6	21.32	40.22	15.08	22.09	26.34
7	20.85	39.94	14.97	21.85	26.04
8	21.04	39.88	14.01	21.67	26.05
9	20.45	39.54	14.72	21.85	25.47
10	20.9	40.09	15.05	22.17	26.13
11	20.54	39.81	14.51	21.59	25.75
12	21.33	40.18	14.85	22.11	26.09
13	20.85	40.04	14.87	22.15	26.17
14	20.56	39.8	15.17	22.08	25.61
15	21.04	40.14	14.44	22.11	26.17
16	21.01	39.99	15.43	22.34	25.88
17	20.94	39.83	14.39	21.57	25.65
18	21.58	40.39	16.28	22.39	26.32
19	21.38	40.12	14.33	22.2	26.14
20	21.44	40.06	14.72	22.32	26.04
21	21.14	40.12	14.96	22.43	26.03
22	21.11	40.11	15.28	22.08	26.25
23	21.04	40.09	14.86	22.19	26.06
24	20.94	39.95	14.78	21.94	25.77
25	20.82	39.96	14.78	22.21	26.03
26	20.61	39.98	15.23	22.4	25.98
27	20.92	39.8	14.77	21.89	25.91
28	20.77	40.05	15.07	22.12	26.13

Tabla 4. Mediciones por variable para la Fase II.

Las especificaciones de ingeniería para las cinco características se muestran en la tabla 5.

Especificaciones de las variables del bimetal					
	Deflexión	Curvatura	Resistividad	Durezab	Durezaa
LSL	19	39	13	20.2	24.5
USL	23	41	17	23.8	27.5
Target	21	40	15	22	26

Tabla 5. Especificaciones para las variables.

3.3. Aplicación de los métodos al caso de estudio.

Se realiza un análisis de Fase I, sobre las variables de la tabla 3, para conocer el estado del proceso. En esta fase se realiza un análisis univariado para conocer el comportamiento de cada una de las variables. Luego, se busca estimar los parámetros (el vector de medias y la matriz de covarianza), verificar la existencia de dependencia entre las variables y verificar la validez del supuesto de normalidad multivariable.

Posteriormente se busca reducir el conjunto de variables mediante el PCA, para finalmente aplicar Cartas de control y determinar Índices de capacidad basados en las componentes principales.

3.3.1. Análisis del estado del proceso en Fase I

Para conocer el estado y la capacidad del proceso en la Fase I o Fase inicial se realizaron los siguientes pasos:

1. Pruebas de normalidad univariada

En primer lugar, se examina la normalidad mediante el uso de dos herramientas estadísticas: histograma y gráfico Q-Q.

El histograma proporciona información visual para establecer una hipótesis inicial de la distribución de los datos, ya que permite identificar la media, la varianza y la asimetría. Además, es posible detectar la existencia de valores atípicos, los cuales se identificarán en forma de colas largas.

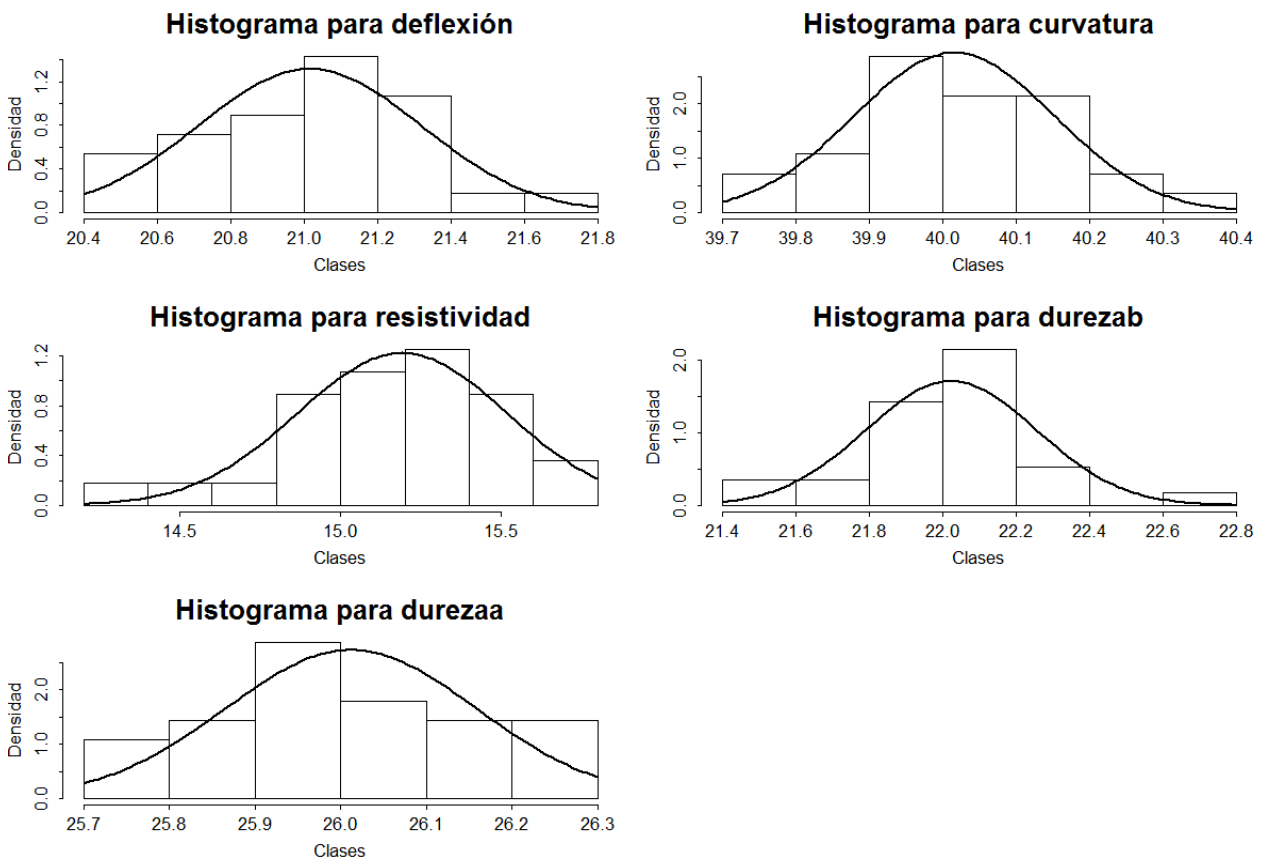


Figura 11. Histograma de las variables individuales.

En la figura 11 se muestran los histogramas para cada una de las variables analizadas. Por el tamaño y ubicación de las clases, se aprecia que la mayoría de los datos se concentran en el centro. Así mismo, al ajustar la curva de una distribución normal sobre cada histograma se puede observar que no existe una asimetría significativa. Agregando, que no se presentan colas largas, es decir no se detectan valores atípicos considerables. Por lo tanto, concluimos que no existe evidencia visual para rechazar nuestra hipótesis inicial.

Los valores exactos para la media, dispersión estándar y varianza se muestran en la tabla 6.

Medidas de tendencia central y dispersión de cada Variable					
	Deflexión	Curvatura	Resistividad	Durezab	Durezaa
Media	21.01607	40.01607	15.19214	22.02393	26.01214
Desviación estándar	0.3031115	0.1361736	0.3260125	0.2333228	0.1465494
Varianza	0.09187659	0.01854325	0.1062841	0.05443955	0.02147672

Tabla 6. Medidas de tendencia central y dispersión para cada una de las variables analizadas.

Por otra parte, el gráfico de cuantiles-cuantiles, o gráfico de Q-Q, permite comparar uno o dos conjuntos de datos con una distribución teórica. Mediante estos gráficos se comparan los cuantiles de cada conjunto de datos contra una línea de referencia a partir de una distribución normal. Se dice que hay evidencia visual que los datos tienen distribución normal si se encuentran sobre la línea de referencia o muy cerca. Con base en lo anterior y a partir de los gráficos Q-Q de la figura 12, cada variable se distribuye normalmente ya que no se presenta desviación significativa de los datos de cada grupo a la línea diagonal.

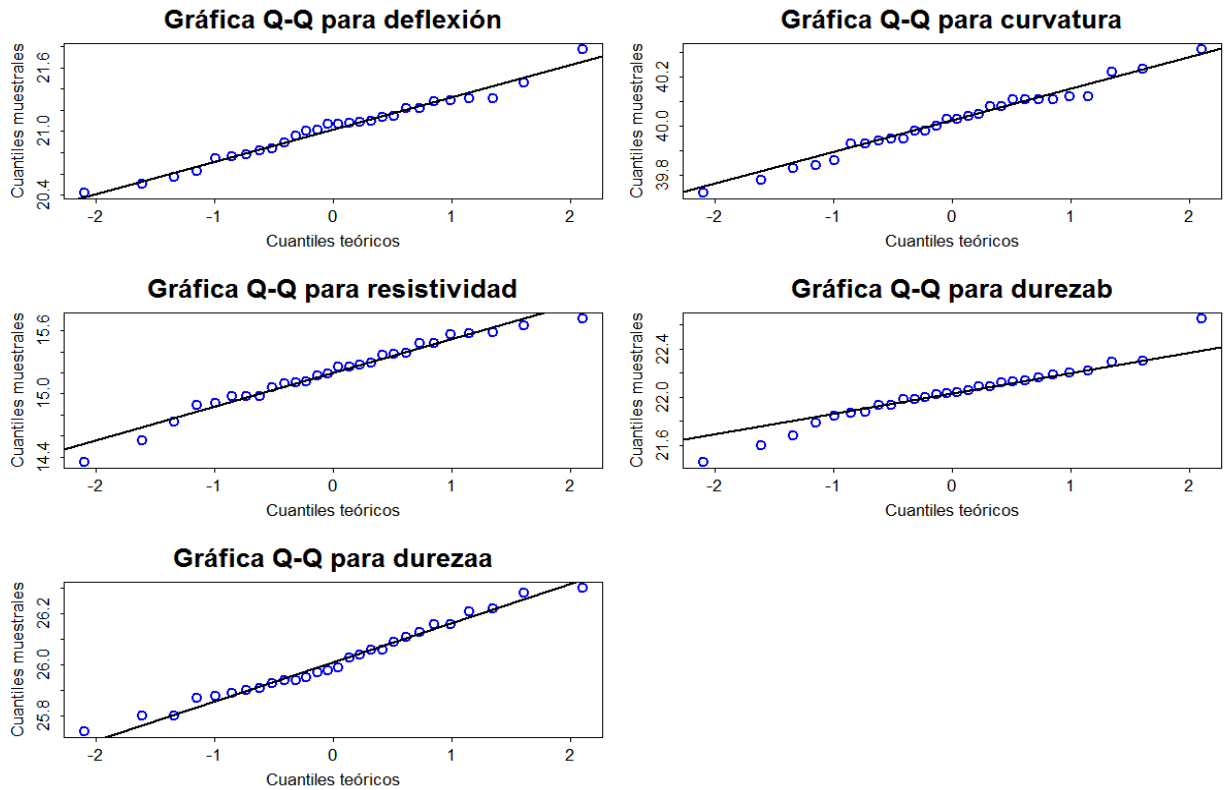


Figura 12. Gráficos Q-Q de las variables individuales.

Existen muchas pruebas de normalidad univariadas, tales como: Anderson-Darling (Anderson y Darling, 1952; Razali y Wah, 2011), Kolmorov-Smirnov (Godina y Matias, 2018; Kolmogorov, 1933), D'Agostino (D'agostino et al., 1990), Jarque-Bera (Carlos M Jarque, 2011; Carlos M Jarque y Bera, 1987), y Shapiro-Wilks (Shapiro y Wilk, 1965) entre otras. Aquí, solo se utiliza y se muestran los resultados de la prueba de Jarque-Bera (consultar en Domanski (2010)). Para que una hipótesis nula no sea rechazada es necesario que el p-valor sea mayor que el nivel de significancia. Recordemos que nuestra hipótesis nula corresponde a:

$$H_0 = \text{Los datos siguen una distribución normal.}$$

De acuerdo con los valores del p-valor generados por la prueba de Jarque-Bera para cada característica de calidad, todos son mayores que 0.05. Por lo tanto, no se encuentra evidencia para rechazar el supuesto de distribución normal a un nivel $\alpha = 0.05$ o $\alpha = 0.10$.

Variable	Chi-Cuadrado	Grados de libertad	p-valor
Deflexión	0.034321	2	0.983
Curvatura	0.09814	2	0.9521
Resistividad	1.7864	2	0.4093
Durezab	1.5836	2	0.453
Durezaa	0.82612	2	0.6616

Tabla 7. Resultados de la prueba de Jarque-Bera para normalidad.

2. Pruebas de normalidad multivariada

Aun cuando los histogramas muestren evidencia de que los datos se distribuyen normalmente para los casos univariados, es importante considerar que la normalidad univariada no necesariamente implica normalidad multivariada. Esto se debe a que, si se presenta una desviación de la normalidad en la distribución de las variables individuales, afectará de manera negativa en la multinormalidad. Por lo tanto, es necesario aplicar otros métodos para probar la normalidad multivariable. Con el objetivo de analizar la normalidad multivariable se aplica la prueba de Mardia (Mardia, 1970). Esta prueba es una generalización de la prueba de asimetría y de curtosis univariada y es una de las más populares en la evaluación de la normalidad multivariada. Los resultados de esta prueba ofrecen un valor de sesgo de 6.982112 con un p-valor igual a 0.585327 mientras que la curtosis es 33.77 con un p-valor de 0.3490892. Luego, dado los p-valores de la asimetría y la curtosis, no hay evidencia para rechazar la normalidad.

En las figuras 13 y 14 se muestra la matriz de dispersión de dos maneras. Esta nos permite crear una visión gráfica sobre la dispersión o la asociación entre las variables. En la figura 14, en diagonal se muestra la distribución de cada variable, mientras que fuera de la diagonal se ubican gráficos de dispersión para cada par de variables. Los pares de variables con mayor dispersión son: Deflexión & Resistividad; Curvatura & Resistividad; Durezab & Deflexión; Durezaa & Deflexión. La matriz de varianzas-covarianzas reafirma conclusión (ver tabla 8).

Matriz de dispersión

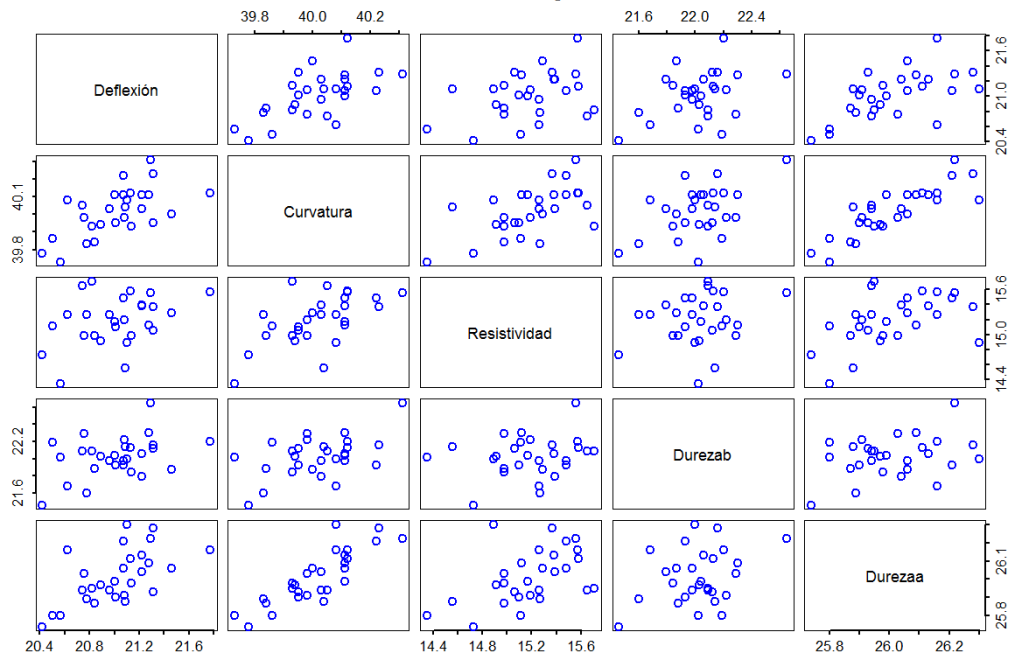


Figura 13. Matriz de dispersión entre las variables.

Matriz de dispersión

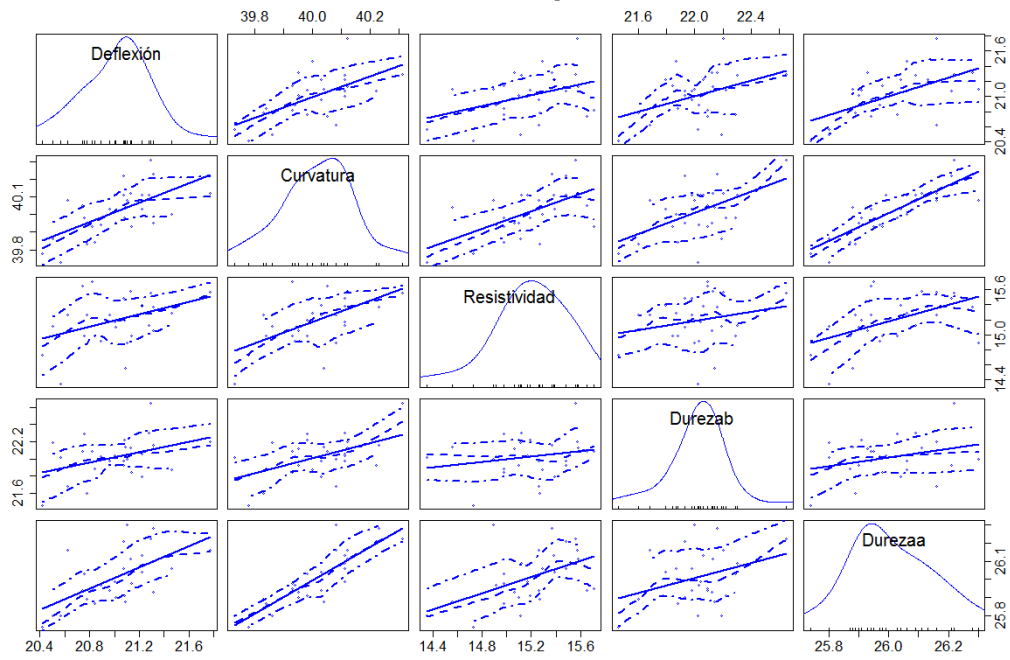


Figura 14. Matriz de dispersión entre las variables.

Matriz de varianzas-covarianzas Fase I					
	Deflexión	Curvatura	Resistividad	Durezab	Durezaa
Deflexión	0.09187659	0.02544325	0.03790873	0.02793082	0.02675317
Curvatura	0.02544325	0.01854325	0.02634206	0.01613082	0.01699762
Resistividad	0.03790873	0.02634206	0.10628413	0.01643942	0.02337672
Durezab	0.02793082	0.01613082	0.01643942	0.05443955	0.01108757
Durezaa	0.02675317	0.01699762	0.02337672	0.01108757	0.02147672

Tabla 8. Matriz de varianzas-covarianzas de las variables.

3. Análisis de Componentes principales

Luego de estudiar el comportamiento de los datos desde el enfoque univariado y el multivariado, el siguiente paso es realizar el análisis de componentes principales.

En las tablas 9 y 10, se presentan los valores y los vectores propios de la matriz de covarianza de los datos originales.

Valores propios de la matriz de datos				
EV1	EV2	EV3	EV4	EV5
0.169984728	0.06588335	0.03964034	0.01484729	0.00226453

Tabla 9. Valores propios de la matriz de covarianzas de las variables.

Vectores propios de la matriz de datos				
Vec1	Vec2	Vec3	Vec4	Vec5
-0.5968081	0.54757583	0.5108459	-0.2880932	-0.00441886
-0.2730603	0.04084787	-0.07387084	0.524618	-0.80192803
-0.6413496	-0.73437024	-0.07519816	-0.2014053	0.05614435
-0.2987809	0.39482181	-0.85052407	-0.1404256	0.10832888
-0.2620282	0.05751741	0.067332	0.7625581	0.58481157

Tabla 10. Vectores propios de la matriz de covarianzas de las variables.

La tabla 11 muestra la desviación estándar para cada componente. Además, se incluye la proporción de varianza y la proporción acumulada.

Componentes principales					
	PC1	PC2	PC3	PC4	PC5
Desviación estándar	0.4123	0.2567	0.1991	0.12185	0.04759
Proporción de varianza	0.5809	0.2251	0.1355	0.05074	0.00774
Proporción acumulada	0.5809	0.8061	0.9415	0.99226	1

Tabla 11. Matriz de varianzas para las componentes principales.

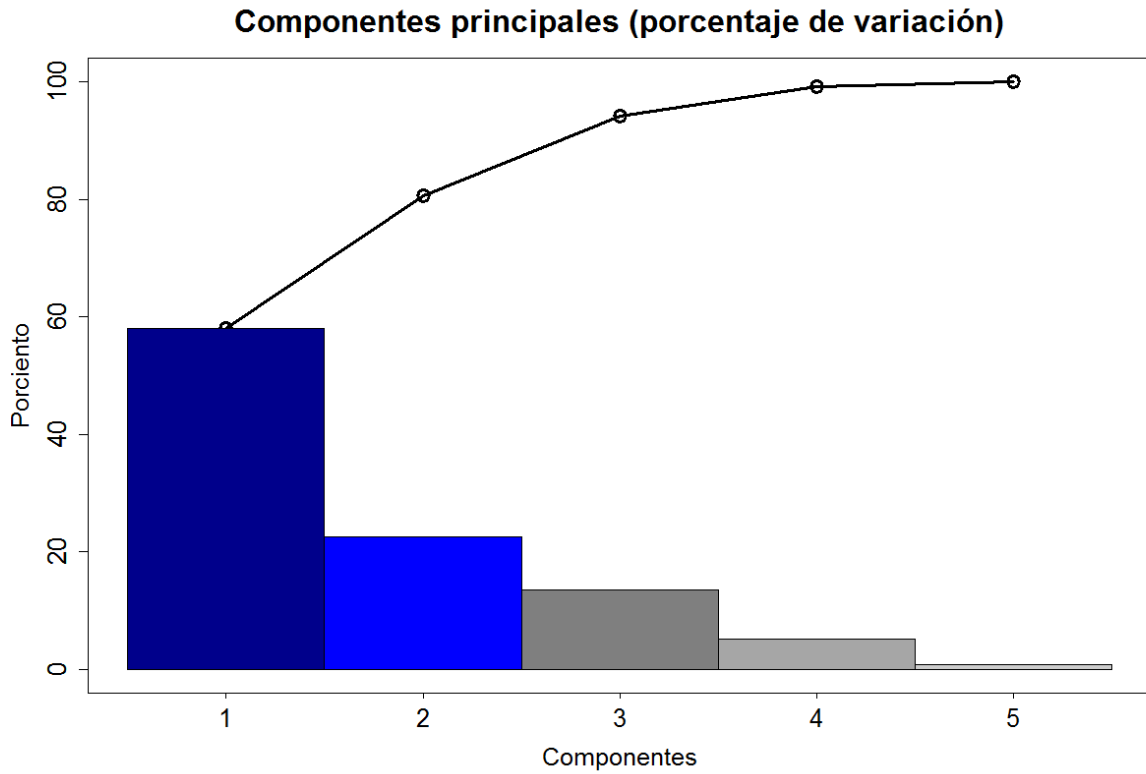


Figura 15. Diagrama de Pareto que muestra el porcentaje de variación de las componentes principales.

En el diagrama de Pareto (figura 15) se aprecia que 80.61% de la variabilidad se concentra en las primeras dos componentes. Por lo tanto, utilizando el método de porcentaje propuesto por Rencher (2002) , el conjunto original se reduce a dimensión 2. Finalmente se incluye la tabla de puntuaciones (tabla 10) para cada componente:

Puntuaciones de los componentes principales					
	PC1	PC2	PC3	PC4	PC5
[1,]	-0.36946526	-0.01281503	0.05185788	-0.08709949	0.03134562
[2,]	-0.28619334	0.13502948	-0.04556845	0.02024799	0.02173226
[3,]	0.40245774	-0.17069315	-0.06227088	0.00331088	0.00664791
[4,]	-0.95639735	-0.23495761	0.20900134	0.0126201	-0.0542425
[5,]	0.72119069	0.15100174	-0.42793303	0.07178157	-0.02688803
[6,]	0.05874719	0.10664423	-0.13815754	-0.14233283	-0.00997024
[7,]	0.45874464	0.10820413	0.02327793	0.1768793	0.00851795
[8,]	0.81201752	0.21761248	0.2092524	-0.15075661	0.04008199
[9,]	0.19506546	-0.46866654	-0.19312556	-0.14963261	0.06977984
[10,]	0.07980828	0.28849754	0.07772795	-0.1688564	0.00663585
[11,]	0.18981113	-0.12382292	0.28910623	-0.03720049	-0.01001232
[12,]	-0.40926326	-0.17527344	-0.40145861	-0.10174913	0.01675338
[13,]	0.29626695	-0.19923851	0.08404912	0.19753972	-0.04207882
[14,]	0.23977974	0.31481442	-0.09624986	0.00837365	-0.00510236
[15,]	0.28986135	0.11254296	0.3547499	-0.09792376	0.0277331
[16,]	-0.35471468	0.54391285	-0.02408453	0.00150597	-0.1197117
[17,]	-0.1382094	0.01810831	0.08104587	-0.08668053	-0.02791956
[18,]	0.24186514	-0.19261829	0.03954891	0.01842675	-0.03617203
[19,]	-0.2394187	-0.3913862	0.09031839	0.29502281	0.00350504
[20,]	-0.0579563	0.2775626	0.10060652	0.29307997	0.09714944
[21,]	0.30949585	-0.00143293	0.06036692	0.03749832	0.00715383
[22,]	0.13898678	-0.46408382	-0.23901917	-0.05917225	-0.03531479
[23,]	-0.21455678	0.12017733	-0.33729346	0.07382929	0.0574139
[24,]	0.00085507	0.01636926	-0.02864356	0.03922433	-0.08770453
[25,]	-0.94120298	0.34874832	-0.15435649	-0.01317561	0.05960909
[26,]	-0.30043107	-0.36844519	0.23963175	-0.07889729	0.03727604
[27,]	-0.01816888	-0.1014601	-0.00227072	-0.03905023	-0.05406086
[28,]	-0.14897553	0.14566807	0.23989076	-0.03681342	0.01784256

Tabla 12. Tabla de puntuaciones de las componentes principales.

4. Gráfico de control T^2

Para evaluar el estado del proceso del Bimetal se realiza el gráfico de control multivariado T^2 de Hotelling. Dicho gráfico se generó utilizando las dos componentes principales obtenidas en el punto anterior.

Previo a la construcción del gráfico de control multivariado T^2 de Hotelling, se generó la elipse de confianza (figura 16) para las dos componentes principales. Se observa que no existen puntos fuera de esta, lo que significa, que no hay evidencia de causas especiales y, por lo tanto, se puede decir que el proceso se encuentra en control.

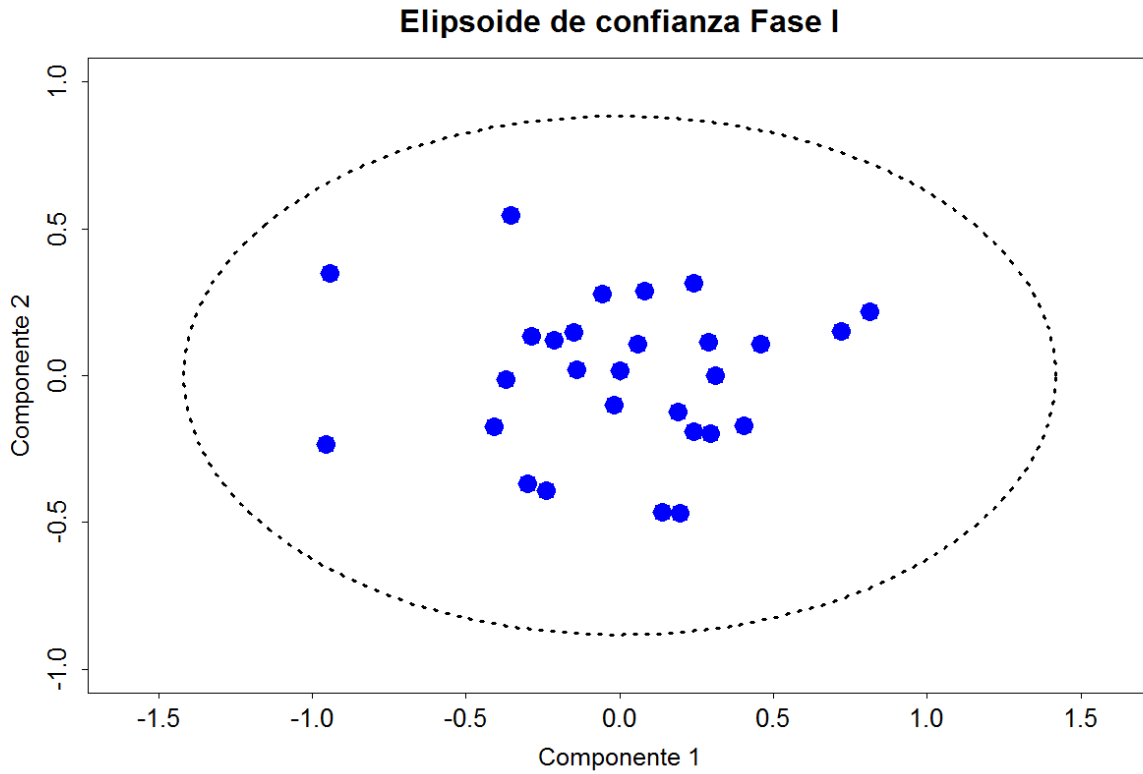


Figura 16. Elipse de confianza para los puntos de las dos componentes principales.

El Gráfico de control T^2 de Hotelling (figura 17) también muestra que todos los puntos se encuentran bajo el límite de control superior.

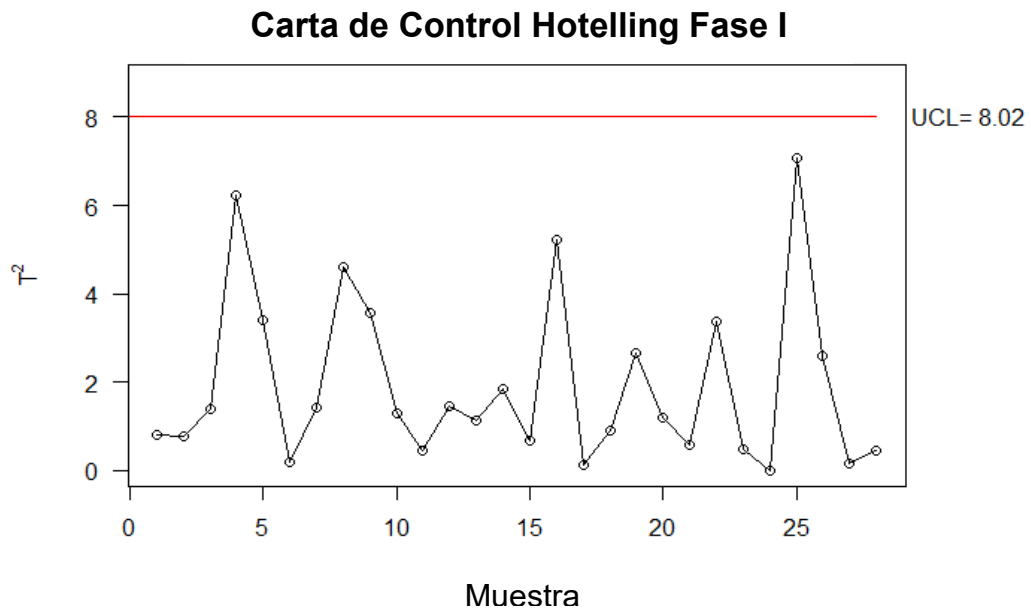


Figura 17. Gráfico de control T^2 de Hotelling para las componentes principales.

Al obtener que todos los puntos de las dos componentes principales, responsables del 80.61% de la variabilidad se encuentran bajo el límite control, se puede concluir que el proceso está bajo control.

5. Índices de capacidad basados en componentes principales.

En esta sección se ilustra muy brevemente el cálculo del índice de Wang y Chen (1998). En la tabla 5, se indicaron los límites de especificación para cada variable analizada, así como los valores objetivos. Retomando esos datos, y recordando que los vectores y los valores propios vienen dados por:

$$U = \begin{pmatrix} -0.5968 & 0.5476 & 0.5108 & -0.2881 & -0.0044 \\ -0.2731 & 0.0408 & -0.0739 & -0.5246 & -0.8019 \\ -0.6413 & -0.7344 & -0.0752 & -0.2014 & -0.0561 \\ -0.2988 & 0.3948 & -0.8505 & -0.1404 & 0.1083 \\ -0.2620 & 0.0575 & 0.0673 & 0.7626 & 0.5848 \end{pmatrix}$$

$$D = (0.1700 \quad 0.0659 \quad 0.0396 \quad 0.0148 \quad 0.0023)$$

Se tiene que, los nuevos vectores de especificaciones vienen dados por;

$$LSL_{PC} = U \cdot LSL = (42.78 \quad 11.83 \quad -9.68 \quad -28.21 \quad 14.11)$$

$$USL_{PC} = U \cdot USL = (50.14 \quad 12.76 \quad -10.95 \quad -29.09 \quad 13.37)$$

$$T_{PC} = U \cdot T = (46 \quad 12 \quad -10 \quad -29 \quad 14)$$

En la tabla 9 se determinó que las primeras dos componentes principales incluían el 80,61% de la variabilidad. Por lo cual se consideró el estudio únicamente de esas dos componentes. Para la primera componente principal, el índice C_p es:

$$C_{p;PC_1} = \frac{USL_{PC_1} - LSL_{PC_1}}{6\sigma_{PC_1}} = \frac{50.14 - 42.78}{6\sqrt{0.1700}} = 2.98$$

Mientras que, para la segunda componente

$$C_{p;PC_2} = \frac{USL_{PC_2} - LSL_{PC_2}}{6\sigma_{PC_2}} = \frac{12.76 - 11.83}{6\sqrt{0.0659}} = 0.60$$

Finalmente, MC_p viene dado por la siguiente expresión: $MC_p = \left(\prod_{i=1}^p C_p; PC_i\right)^{\frac{1}{p}} = (2.98 \cdot 0.60)^{\frac{1}{2}} = 1.34$. Mediante la función `mpci` del Software R determinamos los otros

índices resultando ser iguales a: $MC_{pk} = 1.13$, $MC_{pm} = 1.24$ y $MC_{pmk} = 1.04$. Dado que los índices son superiores a 1, se tiene que el proceso es capaz.

3.3.2. Análisis del estado del proceso en Fase II

Con base en los límites de control en la Fase I, se realiza el análisis de la Fase II mediante una metodología análoga a la de la Fase I.

6. Pruebas de normalidad univariada

En primer lugar, se examina la normalidad mediante el histograma y el gráfico Q-Q. Para cada una de las variables.

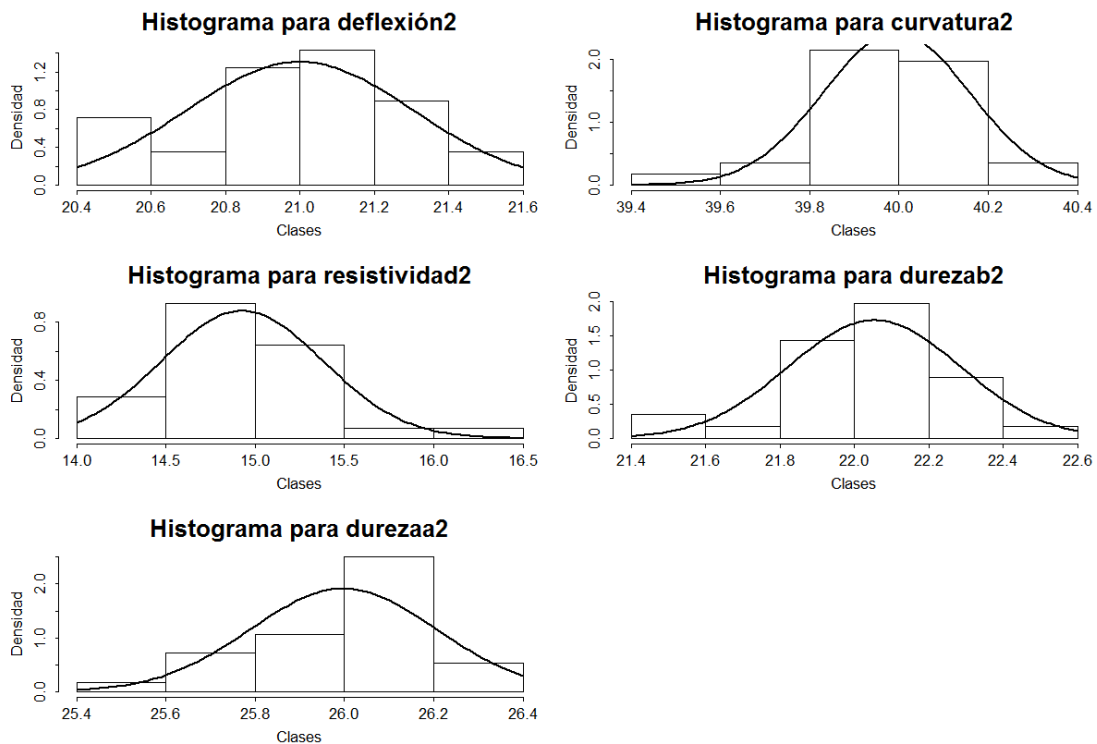


Figura 18. Histograma de las variables individuales para la Fase II.

En la figura 18 se muestran los histogramas correspondientes a cada una de las variables analizadas en la Fase II. En este caso también se observa mayor concentración de los datos en el centro. Así mismo, con el ajuste de la curva de una distribución normal sobre cada histograma se puede notar que no existe una asimetría significativa, ni colas largas. Por lo tanto, se podría decir que las variables se distribuyen normalmente.

Los valores exactos para la media, desviación estándar y varianza se muestran en la tabla 13.

Medidas de tendencia central y dispersión de cada Variable						
	Deflexión	Deflexión	Curvatura	Resistividad	Durezab	Durezaa
Media	21	21	39.99393	14.92714	22.05393	25.9975
Desviación e	0.3050076	0.3050076	0.1646622	0.4537323	0.2314742	0.2084799
Varianza	0.09302963	0.09302963	0.02711362	0.205873	0.05358029	0.04346389

Tabla 13. Medidas de tendencia central y dispersión para cada una de las variables analizadas en la Fase II.

Por otra parte, la gráfica Q-Q para cada una de las variables muestra evidencia de que cada variable se distribuye normalmente ya que no se presenta ninguna desviación significativa de los datos a la línea diagonal (ver figura 19).

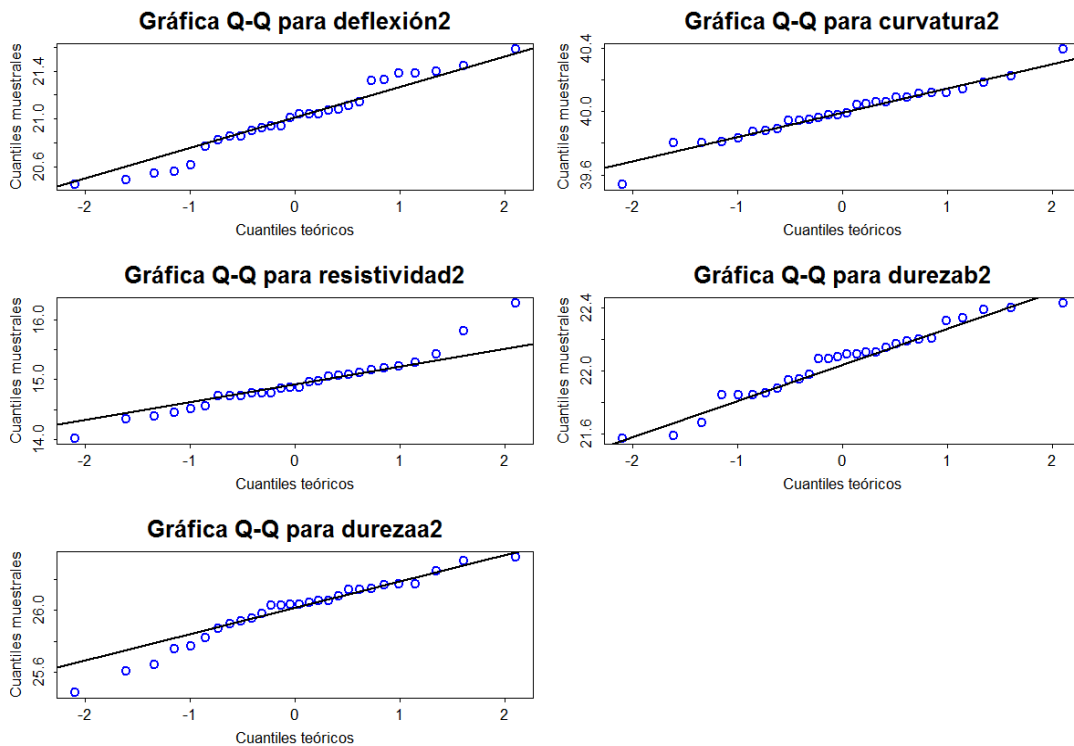


Figura 19. Gráficos Q-Q de las variables individuales para la Fase II.

Al realizar la prueba de Jarque-Bera (ver tabla 14) , se obtuvo que los p-valores para las variables Deflexión, Curvatura, Derezab y Durezaa son mayores que 0.05. Por lo tanto, el supuesto de normalidad no se rechaza para un nivel $\alpha = 0.05$ o $\alpha = 0.10$. Sin embargo, para la variable Resistividad el p- valor es menor que 0.05, con lo cual se rechaza la hipótesis de normalidad para este caso.

Variable	Chi-Cuadrado	Grados de libertad	p-valor
Deflexión	0.64243	2	0.7253
Curvatura	1.6279	2	0.4431
Resistividad	6.2181	2	0.04464
Durezab	0.87399	2	0.646
Durezaa	2.3217	2	0.3132

Tabla 14. Resultados de la prueba de Jarque-Bera para normalidad en la Fase II.

7. Pruebas de normalidad multivariada

Para analizar la normalidad multivariable en la segunda fase, aplicamos nuevamente la prueba de Mardia. En este caso la prueba arroja un valor de sesgo de 6.950879 con un p-valor igual a 0.5924351 mientras que la curtosis es 31.74321 con un p-valor de 0.1515315. Luego, dado los p-valores de la asimetría y la curtosis, no hay evidencia para rechazar la normalidad.

En las figuras 20 y 21 se muestra la matriz de dispersión. Para la fase II, se obtuvo que los pares de variables con mayor dispersión son: Deflexión & Resistividad; Curvatura & Resistividad; Resistividad & Deflexión; Durezab & Deflexión; Durezaa & Deflexión. La matriz de varianzas-covarianzas reafirma esta conclusión (ver tabla 15).

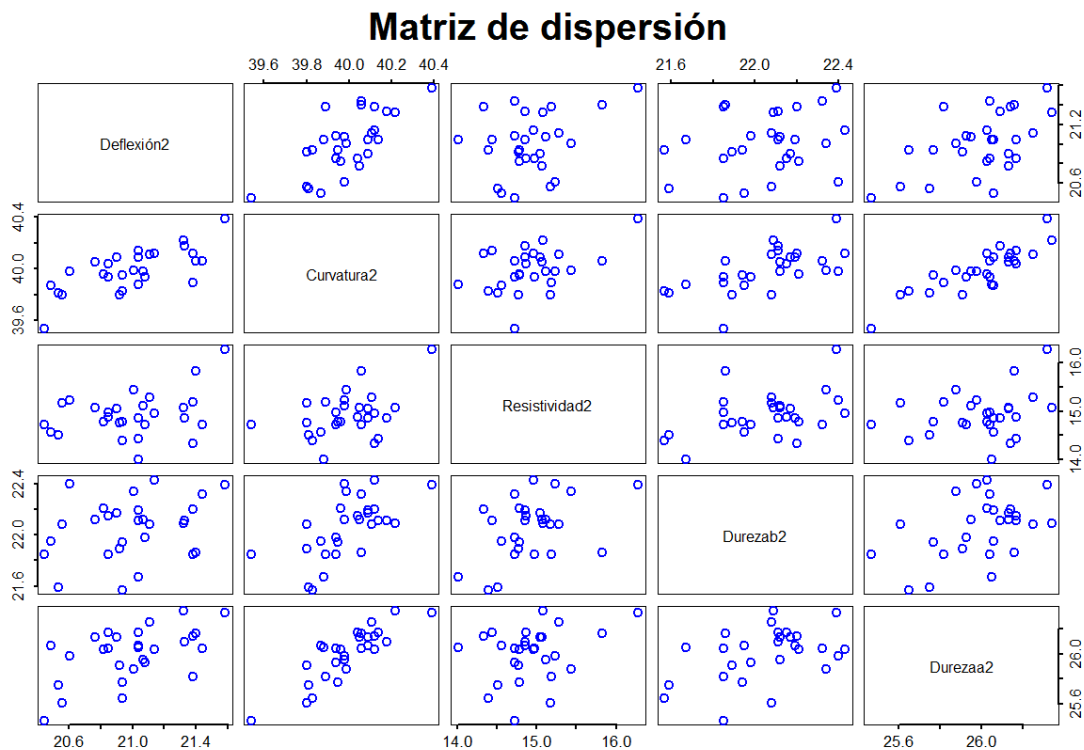


Figura 20. Matriz de dispersión entre las variables para la Fase II.

Matriz de dispersión

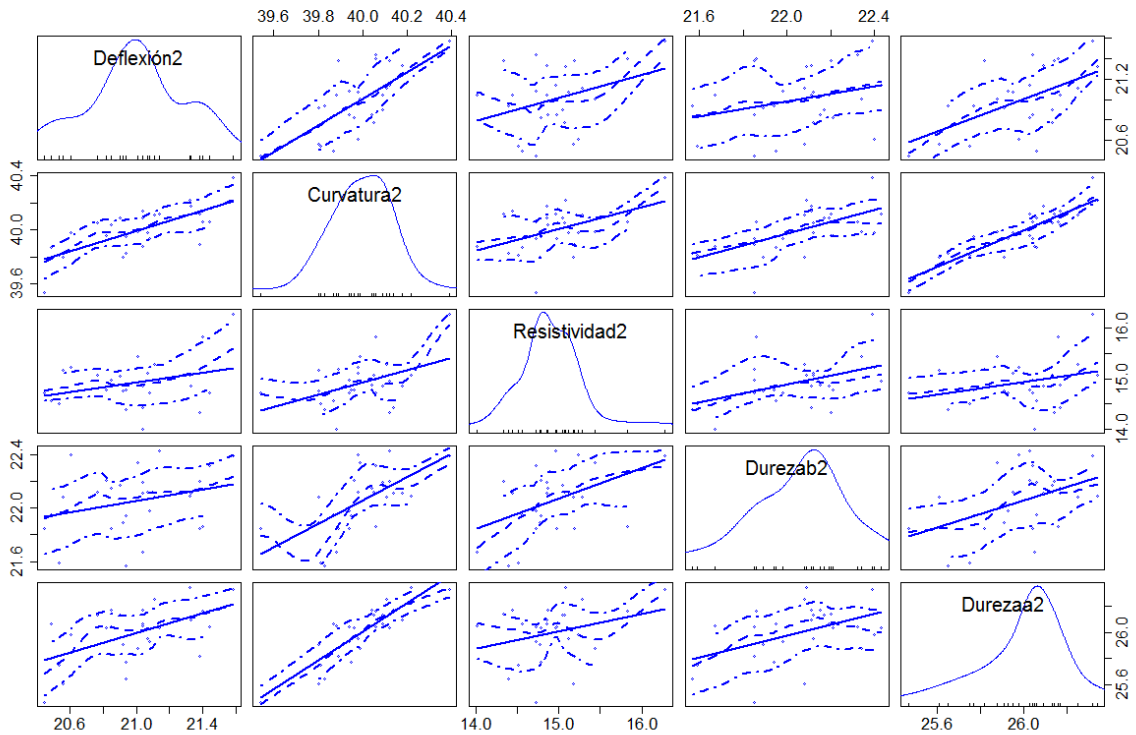


Figura 21. Matriz de dispersión entre las variables para la Fase II.

Matriz de varianzas-covarianzas Fase II					
	Deflexión	Curvatura	Resistividad	Durezab	Durezaa
Deflexión	0.09302963	0.0355037	0.04588519	0.01978889	0.03483704
Curvatura	0.0322037	0.02711362	0.03283757	0.02353585	0.02951759
Resistividad	0.04588519	0.03283757	0.20587302	0.04698201	0.0271037
Durezab	0.01978889	0.02353585	0.04698201	0.05358029	0.02196574
Durezaa	0.03483704	0.02951759	0.0271037	0.0219657	0.04346389

Tabla 15. Matriz de varianzas y covarianzas de las variables en la Fase II.

8. Análisis de Componentes principales

La matriz de componentes (tabla 16) para los datos de la fase II, indica que con dos componentes se puede considerar más del 80% de la variabilidad, específicamente, en este caso corresponde al 83.85%. Información que también se ilustra en el diagrama de Pareto (figura 22).

Componentes principales					
	PC1	PC2	PC3	PC4	PC5
Desviación estándar	0.5078	0.3112	0.211	0.14571	0.05045
Proporción de varianza	0.6096	0.2289	0.1053	0.05019	0.00602
Proporción acumulada	0.6096	0.8385	0.9438	0.99398	1

Tabla 16. Matriz de varianzas para las componentes principales de la Fase II.

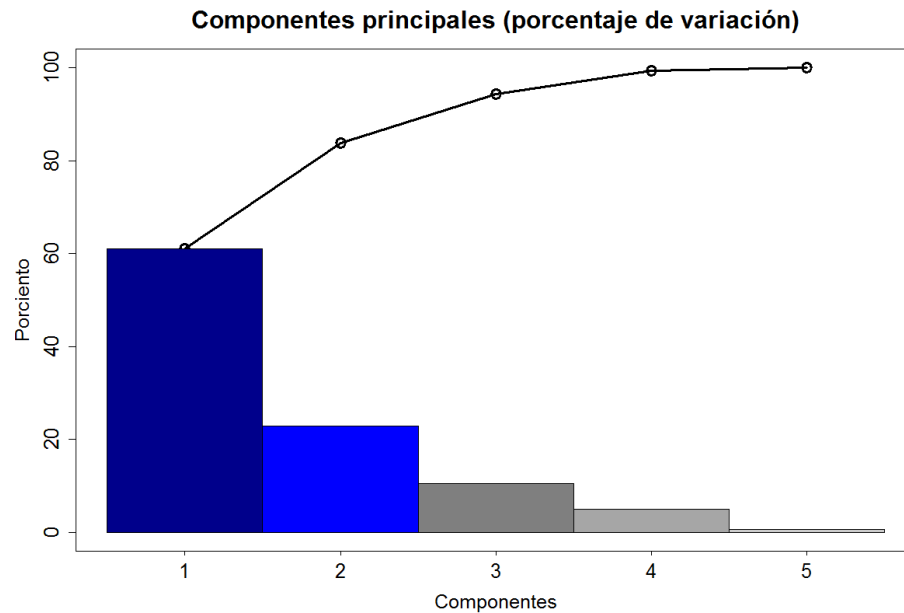


Figura 22. Diagrama de Pareto que muestra el porcentaje de variación de las componentes principales en la Fase II.

9. Gráfico de control T^2

Para evaluar el estado del proceso del Bimetal en la fase II, se analizan los datos de las dos componentes principales, que representan el 83.85% de la variabilidad de las mediciones en esta fase, considerando los límites de control obtenidos en la Fase I. De esta forma, basándonos en el elipsoide de confianza construido en la Fase I, se agregan los datos correspondientes a las componentes principales de la Fase II (representados por las "x" en la figura 23). Se observa entonces, que el dato 18 se encuentra fuera de la elipse de confianza, lo que muestra evidencia de causas especiales y por lo tanto de un proceso que no está en control estadístico.

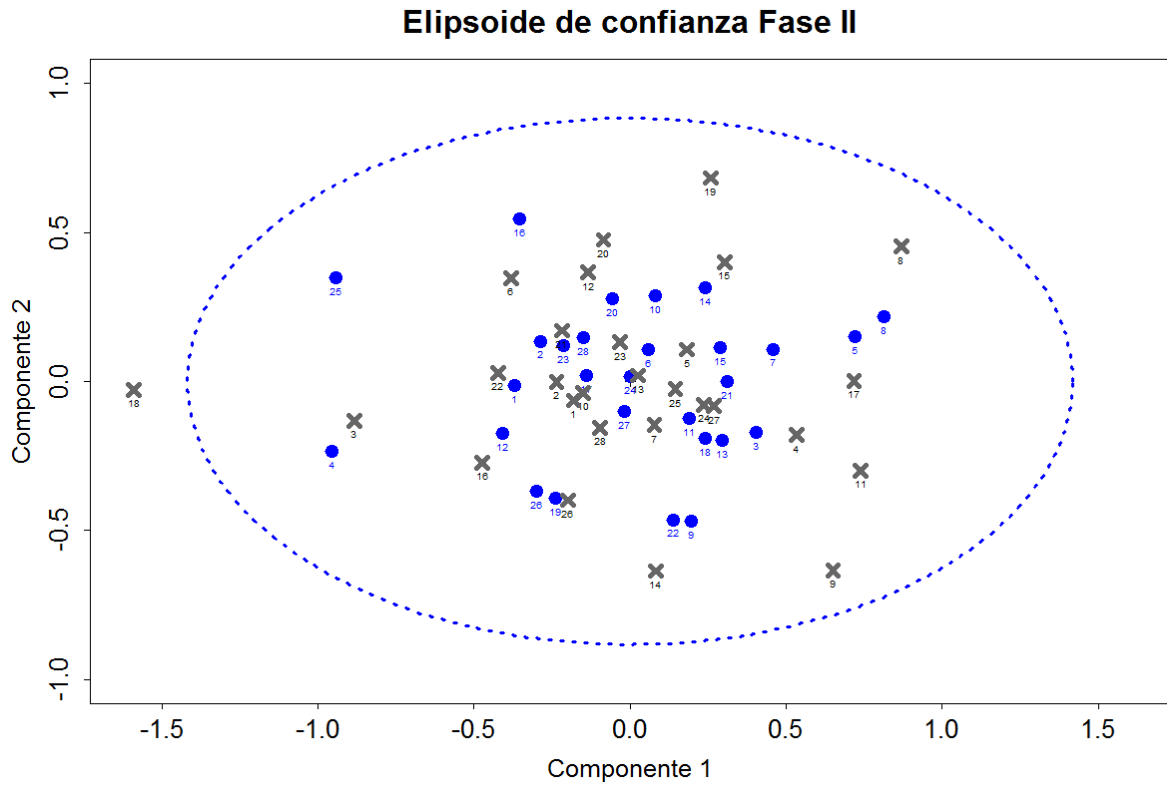


Figura 23. Elipse de confianza para las componentes principales de la Fase II.

En la figura 24 se muestran los gráficos de control T^2 de Hotelling de la Fase I y la Fase II. Se puede notar nuevamente que el dato 18 se encuentra sobre el límite de control superior, con lo cual se confirma que hay evidencia de causas especiales. Por lo tanto, se puede decir que el proceso en la Fase II no se encuentra en control estadístico.

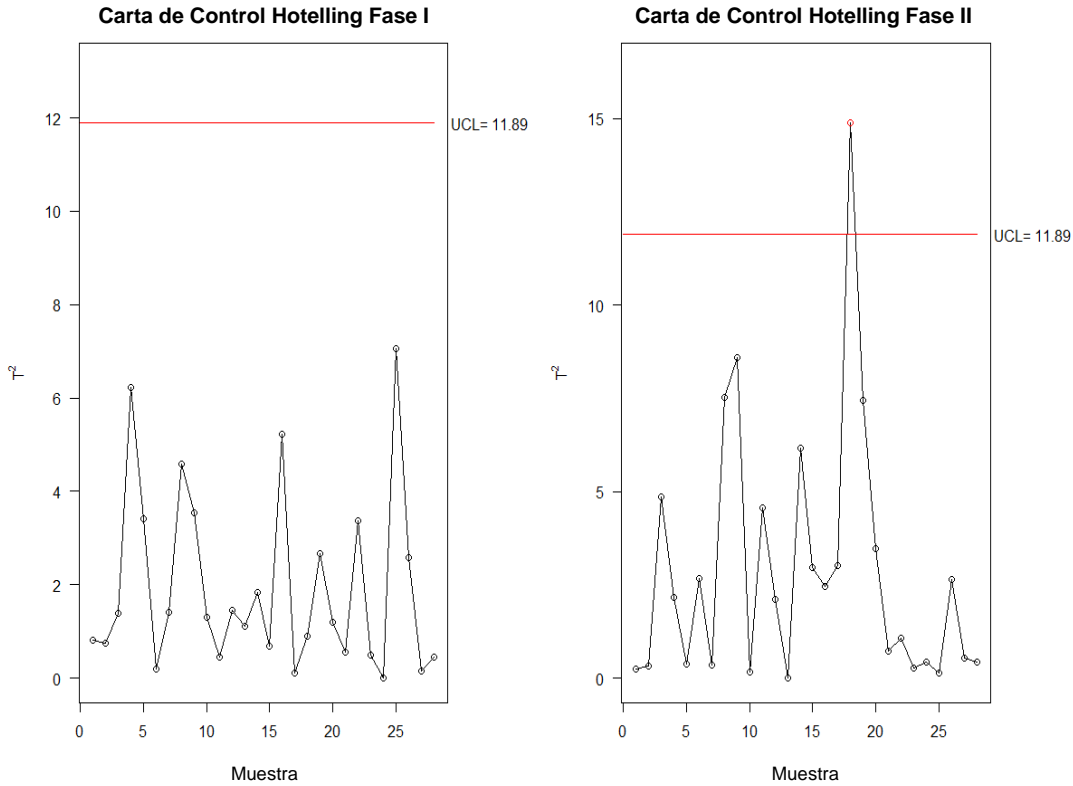


Figura 24. Gráfico de control T2 de Hotelling para las componentes principales de la Fase II.

5. Índices de capacidad basados en componentes principales.

Retomando nuevamente los límites de especificación para cada variable analizada, así como los valores objetivos y, recordando que los vectores y los valores propios vienen dados por:

$$U = \begin{pmatrix} -0.3613 & 0.6833 & 0.5063 & -0.3557 & -0.1406 \\ -0.2306 & 0.3043 & -0.1905 & 0.1988 & 0.8823 \\ -0.8305 & -0.5238 & 0.1684 & 0.0847 & -0.0191 \\ -0.2766 & 0.1088 & -0.7692 & -0.5446 & -0.1532 \\ -0.2237 & 0.3929 & -0.2955 & 0.7282 & -0.4219 \end{pmatrix}$$

$$D = (0.2579 \quad 0.0968 \quad 0.0445 \quad 0.0212 \quad 0.0025)$$

Luego, los nuevos vectores de especificaciones vienen dados por;

$$LSL_{PC} = U \cdot LSL = (37.72 \quad 29.86 \quad -18.40 \quad 8.94 \quad 18.06)$$

$$USL_{PC} = U \cdot USL = (44.62 \quad 32.68 \quad -19.74 \quad 8.47 \quad 17.37)$$

$$T_{PC} = U \cdot T = (45 \quad 33 \quad -20 \quad 8 \quad 17)$$

En la tabla 9 se determinó que las primeras dos componentes principales incluían el 80,61% de la variabilidad. Para la primera componente principal, el índice C_p es:

$$C_{p;PC_1} = \frac{USL_{PC_1} - LSL_{PC_1}}{6\sigma_{PC_1}} = \frac{44.62 - 37.72}{6\sqrt{0.2579}} = 0.58$$

Mientras que, para la segunda componente

$$C_{p;PC_2} = \frac{USL_{PC_2} - LSL_{PC_2}}{6\sigma_{PC_2}} = \frac{32.68 - 29.86}{6\sqrt{0.0968}} = 1.51$$

Finalmente, usando la función “mpci” del Software R determinamos los índices multivariados, resultando ser iguales a: $MC_p = 1.85$, $MC_{pk} = 1.81$, $MC_{pm} = 1.84$ y $MC_{pmk} = 1.80$. Dado que los índices son superiores a 1, se tiene que el proceso es capaz.

Capítulo 4. Discusión y análisis de resultados

4.1. Resultados

El estudio se realizó sobre 5 características de calidad medidas en un proceso de producción de Bimetal para termostato. Dichas características deben cumplir ciertos límites de especificación establecidos por el área de Ingeniería e incluidos en la tabla 5. Para determinar el estado y la capacidad del proceso, se realizó un análisis de componentes principales, y sobre las componentes que representaron más del 80% de la variabilidad, se aplicó control estadístico multivariado de procesos mediante la implementación de los gráficos T2 de Hotelling. Además, se calcularon los índices de capacidad multivariados basados en PCA.

Para la Fase I, debido a que no se detectaron puntos fuera de los límites de control en la elipse de confianza ni en el gráfico de control T2, lo que resultó en no tener evidencia de causas especiales y, en consecuencia, se determinó que el proceso estuvo bajo control. En caso contrario, en la elipse de confianza y el gráfico de control T2 para los datos de la Fase II, se detectó un punto por fuera de los límites de control, lo que indicó la presencia de causas especiales en el proceso. Luego, se concluyó que el proceso no estaba bajo control en la Fase II. Por otro lado, los índices de capacidad de la Fase

Indicaron que la capacidad potencial del proceso fue adecuada, al obtener un índice MC_p mayor a 1.33. Mientras que la capacidad real del proceso resulto parcialmente adecuada ya que los índices MC_{pk} y MC_{pm} tuvieron valores entre 1 y 1.33. Respecto a la Fase II, la capacidad potencial y real son adecuadas, ya que los índices de capacidad MC_p , MC_{pk} y MC_{pm} fueron mayores que 1.33.

4.2. Importancia o trascendencia.

La aplicación del análisis de componentes en las dos fases, contribuyo de manera positiva a la generación de los gráficos de control y al cálculo de los índices de capacidad. Pues el problema de trabajar con 5 variables se redujo considerablemente al trabajar solo con dos componentes que representaron más del 80% de la variabilidad del proceso, pasando así, a un caso de análisis bivariado.

Los resultados que se obtuvieron de la implementación del MSPC, permiten al grupo de Control de calidad investigar cuales con las causas especiales que puedan estar contribuyendo a que el proceso no se encuentre bajo control, para poder eliminarlas. Además, por los resultados de los índices de capacidad, al obtener un proceso con capacidad aceptable en la Fase II se debe poner atención en las acciones de mejora que se implementaron durante el periodo de tiempo entre la Fase I y la Fase II, para controlar que continúen e integrar nuevas acciones que puedan contribuir a un incremento aún mayor de la capacidad del proceso.

Conclusiones y recomendaciones

La implementación adecuada del control estadístico de procesos genera importantes mejoras en la detección de causas especiales de variabilidad, y como consecuencia para el incremento de la calidad de los productos y procesos. Además, las normas del Sistemas de Gestión de la Calidad establecen como requisito que las empresas realicen el control de sus procesos oportuna y adecuadamente. Por lo tanto, el control estadístico de procesos debería aplicarse de forma univariada cuando así sea suficiente, pero no excluirse el caso multivariado cuando sea necesario por la presencia de un número muy grande de variables, situación en la cual, no es necesario trabajar con todo el conjunto de variables sino solo una reducción aplicando PCA.

El uso de métodos como el PCA, resulta de bastante utilidad para aplicar las técnicas multivariadas de control estadístico de procesos. Debido a que, si la correlación entre las variables es alta, se podría generar pérdida de información significativa para conocer el estado del proceso. Por otro lado, bajo las mismas condiciones, el conjunto de las componentes, que proporcionen la mayor información sobre la variación del proceso podría reducirse bastante y ser más fácil de analizar.

La realización de esta investigación deja camino para seguir aplicando otros métodos que pueden generar un análisis más completo, y que por limitación de recursos no se realizó. Una de las aplicaciones podría ser:

- Usar alguno de los métodos de descomposición para saber con mayor exactitud, cual o cuales de las características de calidad están generando las causas especiales dentro del proceso de producción del Bimetal. El método más aceptado es el de descomponer la estadística T2 en partes independientes propuesto por Masson (consultar Bersimis, Psarakis y Panaretos (2007) o Masson (1997)) .

Anexo

Código implementado en el Software R para generar los gráficos y resultados estadísticos.

```
library("MPCI")
library("MSQC")
library("tseries")
##### FASE I #####
data(bimetal1)
cols <- c("deflexión", "curvatura", "resistividad", "durezab", "durezaa")
colnames (bimetal1) <- cols
str (bimetal1)
#Análisis de la normalidad univariada.
  par(mfrow=c(3,2))
  for(i in 1:ncol(bimetal1))
  {
    x <-bimetal1[,i]
    mean <-mean(bimetal1[,i])
    sd<-sd(bimetal1[,i])
    hist(x, prob=TRUE, main=paste("Histograma para", colnames(bimetal1)[i]),
    xlab="Clases", ylab="Densidad",
    cex.main=2.5, cex.lab=1.65, cex.axis=1.6, cex=2.1, tcl=-0.3)
    #Curva normal
    points(curve(dnorm( x, mean=mean, sd=sd), add=TRUE), type="l", lwd=2)
  } //Con este ciclo se generan los histogramas y la curva normal ajustada, para
  cada una de las variables.
# Gráfico Q-Q para cada variable:
  par(mfrow=c(3,2))
  for(i in 1:ncol(bimetal1))
  {
    qqnorm(bimetal1[,i], main=paste("Gráfica Q-Q para", colnames(bimetal1)[i]),
    xlab="Cuantiles teóricos", ylab="Cuantiles muestrales",
```



```
cex.main=2.5, cex.lab=1.65, cex.axis=1.6, cex=2.1, tcl=0.3, col="blue",  
bg="blue", lwd=2.5)
```

```
#Línea de referencia de la distribución normal
```

```
qqline(bimetal1[,i], lwd=2)
```

```
}
```

```
# Prueba de Jarque-Bera de normalidad univariada
```

```
jarque.bera.test(bimetal1[,1])
```

```
jarque.bera.test(bimetal1[,2])
```

```
jarque.bera.test(bimetal1[,3])
```

```
jarque.bera.test(bimetal1[,4])
```

```
jarque.bera.test(bimetal1[,5])
```

Análisis de la normalidad multivariada

```
# Prueba de Mardia de normalidad multivariada
```

```
MardiaTest(bimetal1)
```

```
#Matriz de gráficos de dispersión para cada par de variables
```

```
scatterplotMatrix (bimetal1, main=paste("Matriz de dispersión"),
```

```
cex.main=2.5, cex.lab=1.9, cex.axis=1.6)
```

```
pairs(bimetal1,labels=c("Deflexión", "Curvatura", "Cesistividad", "Durezab",  
"Durezaa"), main=paste("Matriz de dispersión"), cex.main=2.5, cex.lab=1.9,  
cex.axis=1.6, cex=2.1, tcl=-0.3,col="blue", lwd=2.5 )
```

Análisis de componentes

```
eigen(covariance(bimetal1))
```

```
summary(prcomp(bimetal1))
```

```
varian <-(prcomp(bimetal1)$sdev) ^ 2
```

```
perc <-varian / sum(varian)
```

```
cumperc <-cumsum(perc)
```

```
cumperc100<-cumperc*100
```

```
perc100<-perc*100
```

```
par(mfrow=c(1,1))
```

```
plot(cumperc100, main=paste("Componentes principales (porcentaje de  
variación)"), type = "o", xlim = c(0.5, length(cumperc) + 0.5), ylim = c(0,1),
```

```

xlab = "Componentes", ylab = "Porciento", lwd=3, cex.main=2.0, cex.lab=1.55,
cex.axis=1.6, cex=1.8,tcl=-0.3 )
barplot (perc100, add = TRUE, width = 1, beside = TRUE, col=c("blue4", "blue",
"gray50", "gray65", "gray80"), space = c(0, 0.5), cex.axis=1.6)
S <-covariance(bimetal1)

```

#Grafico elipsoidal basada en componentes principales

```

a <-predict(prcomp(bimetal1))[,1:2]
S <-covariance(a)
Xmv <- colMeans(a)
plot(ellip(type ="t2", a, alpha=0.01), main=paste("Elipse de confianza Fase I"),
type="l",lty = 3, xlim=c(- 1.6,1.6), ylim=c(- 1,1), xlab = "Componente 1", ylab =
"Componente 2", cex.main=2.0, cex.lab=1.55, cex.axis=1.6, tcl=-0.3, lwd=3)
points(Xmv [1], Xmv [2], pch = 3)
points(a, cex = 0.75, col="blue", cex=1.5, lwd=10)

```

#Carta de control basada en componentes principales

```

par(mfrow = c(1,1))
mult.chart(a, type = "t2", Xmv=Xmv, S=S, alpha = 0.01)

```

FASE II

```

data(bimetal2)
colsb <- c("deflexiónb", "curvaturab", "resistividadb", "durezabb", "durezaab")
colnames(bimetal2) <- colsb
str(bimetal2)

```

#Análisis de la normalidad univariada.

```

par(mfrow=c(3, 2))
for(i in 1:ncol(bimetal2))
{
x <-bimetal2[,i]
mean <-mean(bimetal2[,i])
sd<-sd(bimetal2[,i])
hist(x,prob=TRUE,main=paste("Histograma para", colnames(bimetal2)[i]),

```

```

xlab="Clases", ylab="Densidad", cex.main=2.5, cex.lab=1.65, cex.axis=1.6,
cex=2.1, tcl=-0.3)
#Curva normal
points(curve(dnorm(x, mean=mean, sd=sd), add=TRUE), type="l", lwd=2)
} //Con este ciclo se generan los histogramas y la curva normal ajustada, para
cada una de las variables.
# Gráfico Q-Q para cada variable:
par(mfrow=c(3, 2))
for(i in 1:ncol(bimetal2))
{
qqnorm(bimetal2[,i],main=paste("Gráfica Q-Q para", colnames(bimetal2)[i]),
xlab="Cuantiles teóricos", ylab="Cuantiles muestrales", cex.main=2.5,
cex.lab=1.65, cex.axis=1.6, cex=2.1, tcl=-0.3, col="blue", bg="blue", lwd=2.5)
#Línea de referencia de la distribución normal
qqline(bimetal2[,i], lwd=2)
}
# Prueba de Jarque-Bera de normalidad univariada
jarque.bera.test(bimetal2[,1])
jarque.bera.test(bimetal2[,2])
jarque.bera.test(bimetal2[,3])
jarque.bera.test(bimetal2[,4])
jarque.bera.test(bimetal2[,5])
# Análisis de la normalidad multivariada
# Prueba de Mardia de normalidad multivariada
MardiaTest(bimetal2)
#Matriz de gráficos de dispersión para cada par de variables
scatterplotMatrix(bimetal2, main=paste("Matriz de dispersión"),
cex.main=2.5, cex.lab=1.9, cex.axis=1.6)
pairs(bimetal2, labels=c("Deflexión2", "Curvatura2", "Resistividad2",
"Durezab2", "Durezaa2"), main=paste("Matriz de dispersión"), cex.main=2.5,
cex.lab=1.9, cex.axis=1.6, cex=2.1, tcl=-0.3, col="blue", lwd=2.5)

```

Análisis de componentes

```
eigen(covariance(bimetal2))
summary(prcomp(bimetal2))
varian <-(prcomp(bimetal2)$sdev) ^ 2
perc <-varian / sum(varian)
cumperc <-cumsum(perc)
cumperc100<-cumperc*100
perc100<-perc*100
par(mfrow=c(1, 1))
plot(cumperc100, main=paste("Componentes principales (porcentaje de
variación)", type = "o", xlim = c(0.5, length(cumperc) + 0.5), ylim = c(0,1),
xlab = "Componentes", ylab = "Porcentaje", lwd=3, cex.main=2.0, cex.lab=1.55,
cex.axis=1.6, cex=1.8, tcl=-0.3)
barplot(perc100, add = TRUE, width = 1, beside = TRUE, col=c("blue4", "blue",
"gray50", "gray65", "gray80"), space= c(0, 0.5), cex.axis=1.6)
```

#Grafico elipsoidal basada en componentes principales

```
#Elipsoide de la fase I
a <-predict(prcomp(bimetal1))[,1:2]
S <-covariance(a)
Xmv <- colMeans(a)
plot(ellip(type ="t2", a, alpha=0.01), main=paste("Elipsoide de confianza Fase
II"), type="l", lty = 3, xlim=c(- 1.6,1.6), ylim=c(-1,1), xlab = "Componente 1", ylab
= "Componente 2", cex.main=2.0, cex.lab=1.55, cex.axis=1.6, tcl=-0.3, lwd=3,
col="blue")
points(Xmv [1], Xmv [2], pch = 3)
points(a, col="blue",cex=1.0,lwd=8,text(a[,1], a[, 2],col="blue", cex=0.6, pos=1,
offset=0.5))
#Elipsoide de control para la fase II
b <-cbind(predict(prcomp(bimetal2))[,1 : 2], 1 : nrow(bimetal1))
points(b[, 1], b[, 2], pch = 4,
cex = 1.5, col="gray40", lwd=5,
```

```
text(b[,1], b[, 2],labels= b[, 3], cex=0.6, pos=1, offset=0.5))
```

#Carta de control basada en componentes principales

```
Xmv2 <- mult.chart(a, type = "t2") $Xmv
```

```
S2 <- mult.chart(a, type = "t2") $covariance
```

```
par(mfrow = c(1,2))
```

```
mult.chart(a, type = "t2", Xmv=Xmv, S=S, alpha = 0.01)
```

```
mult.chart(b[,1:2], type = "t2", Xmv=Xmv2, S=S2, alpha = 0.01)
```

Cálculo de los índices de capacidad multivariados

```
x <- bimetal1
```

```
y<-bimetal2
```

```
Target <- c(21, 40, 15, 22, 26)
```

```
LSL <- c(19, 39, 13, 20.2, 24.5)
```

```
USL <- c(23, 41, 17, 23.8, 27.5)
```

```
#Fase I
```

```
mpci(index = "wang", x, LSL, USL, Target, method = 1, perc = 0.80)
```

```
#Fase II
```

```
mpci(index = "wang", y, LSL, USL, Target, method = 1, perc = 0.80)
```

Referencias

- American Society for Quality. (2017). Recuperado el 8 de mayo de 2019 de <https://asqlatam.org/>
- Anchiraico A., W. R. (2003). *Métodos multivariantes en control estadístico de la calidad* (Trabajo monográfico de pregrado). Universidad Nacional Mayor de San Marcos. Lima, Perú.
- Argumedo, O. J., Molina, R. D., Martínez, E. A. y Hernández, J. A. (2017). Control estadístico multivariante de proceso aplicado en la industria. *CULCyT*, 14(63), 102-116.
- Asociación Mexicana de la Industria Automotriz, A. C. (2011). Recuperado el 24 de abril de 2019 de <http://www.amia.com.mx/ventasp.html>
- Bartlett, M. S. (1954). A note on the multiplying factors for various X² approximations. 296-298.
- Bersimis, S., Psarakis, S., y Panaretos, J. (2007). Multivariate statistical process control charts: an overview. *Quality and reliability engineering international*, 23 (5), 517-543.
- Camacho, J., Pérez- Villegas, A., García- Teodoro, P., Maciá-Fernández, G. (2016). PCA- based multivariate statistical network monitoring for anomaly detection. *Computers & security*, 59, 118-137.
- Carro, P. R. y González, G. D. (2012). *Administración de las Operaciones. Control estadístico de procesos*. Universidad nacional de mar de la plata.
- Centro de Información América Latina. (2019). Recuperada el 25 de abril de 2019 de https://www.cialdnb.com/pdf/economic-analysis/food-and-beverages/MEX Economic-Analysis_ES.pdf //
- Chakraborty, A. K., y Chatterjee, M. (2016). Univariate and Multivariate Process Capability Analysis for Different Types of Specification Limits. *Quality and Reliability Management and Its Applications* (pp. 47-81). London: Springer London.

- Chan, L. K., Cheng, S. W., Spiring, F. (1991). A multivariate measure of process Capability. *International Journal of Modelling and Simulation*, 11(1), 1-6.
- Chang, L. K., Cheng, S. W. y Spiring, F. A. (1988). A New Measure of Process Capability: Cpm. *Journal of Quality Technology*, 20(3), 162-175.
- Chen, H. (1994). A multivariate process capability index over a rectangular solid tolerance zone. *Statistica Sinica* 4 (2), 749-758.
- Confederación de Asociaciones de Agentes aduanales de la República Mexicana. (2018). Recuperado el 25 de abril de 2019 de <http://caaarem.mx/2019/01/11/mexico-se-mantiene-como-cuarto-exportador-de-vehiculos-a-nivel-mundial/>
- Das, N., y Prakash, V. (2008). Interpreting the out-of-control signal in multivariate control chart- a comparative study. *The International Journal of Advanced Manufacturing Technology*, 37(9-10), 966-979.
- De-Felipe, D., y Benedito, E. (2017 a). Monitoring high complex production processes using process capability indices. *The International Journal of Advanced Manufacturing Technology*, 93(1-4), 1257-1267.
- De-Felipe, D., y Benedito, E. (2017 b). A review of univariate and multivariate process capability indices. *The International Journal of Advanced Manufacturing Technology*, 92(5-8), 1687-1705. Doi:10.1007/s00170-017-0273-6.
- Domanski, C. (2010). *Properties of the Jarque-Bera test*, 235.
- Escobar, A. (2019). *México se mantiene como cuarto exportador de vehículos a nivel mundial*. Forbes México. Recuperado de <https://www.forbes.com.mx/mexico-se-mantiene-como-cuarto-exportador-de-vehiculos-a-nivel-mundial/>
- Gutiérrez, P. H., y de la Vara, S. R. (2013). *Control estadístico de la calidad y seis sigma*. (3ra. ed.). México, D.F.: Mc Graw Hill.
- Hadian, H., y Rahimifard, A. (2019). Multivariate statistical control chart and process capability indices for simultaneous monitoring of project duration and cost. *Computers & Industrial Engineering*, 130, 788-797.
- Herrera, A. R., Hernández, A. K., Figueroa, G. E. y De La Ossa, D. Á. J. (2018). Aplicación del control estadístico multivariado para medir la capacidad del

proceso de fabricación de resortes de compresión en acero inoxidable. *Prospectiva*, 16(2), 48-58.

- Hotelling, H. J. (1947). Multivariate quality control. Techniques of statistical analysis.
- Hubele, N., Shahriari, H., y Cheng, C. (1991). A bivariate process Capability vector.
- Instituto Nacional de Estadística y Geografía. (2019). *Comunicado de prensa* (143/19). Recuperado de https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2019/pib_pconst/pib_pconst2019_02.pdf
- Joshi, K. N., y Patil, B. T. (2019). A perspective of Integrated Machine Vision Based-Multivariate Statistical Process Control. *Proceedings of International Conference on Intelligent Manufacturing and Automation*, 463-471.
- Kim, S. B., Sukchotrat, T., y Park, S. K. (2011). A nonparametric fault isolation approach through one-class classification algorithms. *IIE Transactions*, 43(7), 505-517.
- López, R. L. (2016). *Implantación de un sistema de calidad: los diferentes sistemas de calidad existentes en la organización*. (1ra. ed.). España: Ideas propias.
- Lowry, C. A., y Montgomery, D. C. (1995). A review of multivariate control charts. *IIE Transactions*, 27(6), 800-810.
- MacGregor, J. F., y Kourti, T. (1995). Statistical process control of multivariate processes. *Control Engineering Practice*, 3(3), 403-414.
- Mason, R. L., Champ, C.W., Tracy, N.D., Wierda, S.J., Young, J. C. (1997). Assessment of multivariate process control techniques. *Journal of Quality Technology*, 29(2), 140-143.
- Mason, R. L., y Young, J. C. (2002). *Multivariate statistical process control with industrial applications*. Philadelphia, Pennsylvania: Siam.
- Montgomery, D. C. (2009). *Introduction to statistical quality control*. (6ta. ed.) Arizona, United States of America: WILEY.

- Mosquera, J., Olaya, J. y Escobar, R. (2007). Aplicación del control estadístico multivariante en un proceso de extrusión de película plástica. *Scientia et Technica*, (37), 333-338.
- Organización Internacional de Constructores de Automóviles. (2019). Recuperado el 24 de abril de 2019 de <http://www.oica.net/category/about-us/organizations/>
- Pan, J. N., y Lee, C. Y. (2010). New capability indices for evaluating the performance of multivariate manufacturing processes. *Quality and Reliability Engineering International*, 26(1), 3-15.
- Park S. H., y Kim S. B. (2018). Multivariate control charts that combine the Hotelling T2 and classification algorithms. *Journal of the Operational Research Society*, 1-9.
- Pearn, W. L., Kotz, S., y Johnson, N. L. (1992). Distribution and Inferential Properties of Process Capability Indices. *Journal of Quality Technology*, 24(4), 216-231.
- Phaladiganon, P., Kim, S. B., Chen, V. C., y Jiang, W. (2013). Principal component analysis-based control charts for multivariate nonnormal distributions. *Expert Systems with Applications*, 40(8), 3044-3054.
- Praus, P. (2018). Evaluation of biological wastewater treatment process using Mahalanobis distances in original and principal component space: a case study. *Applied Water Science*, 8(6), 167. doi:10.1007/s13201-018-0794-7
- Rencher, A. C. (2003). *Methods of multivariate analysis*. (2da. ed.). Canada: John Wiley & Sons, WILEY.
- Rogalewicz, M. (2012). Some notes on multivariate statistical process control. *Management and Production Engineering Review*, 3(4), 80-86. Doi: 10.2478/v10270-012-0036-7.
- Rogalewicz, M., y Poznańska, P. (2013). The Methodology of Controlling Manufacturing Processes with the Use of Multivariate Statistical Process Control Tools. *Trends in the Development of Machinery and Associated Technology*, 17(1), 89-93.

- Santos-Fernández, E. (2012). *Multivariate statistical quality control using R*. (Vol. 14). Santa Clara, Cuba: Springer Science & Business Media.
- Shahriari, H., Hubele, N., & Lawrence, F. (1995). A multivariate process capability vector.
- Taam, W., Subbaiah, P., & Liddy, J. W. (1993). A note on multivariate capability indices. *Journal of Applied Statistics* 20(3), 339-351.
- Tong, Y. L. (2012). *The multivariate normal distribution*: Springer Science & Business Media.
- Tracy, N. D., Young, J. C., & Mason, R. L. (1992). Multivariate control charts for individual observations. *Journal of Quality Technology*, 24(2), 88-95.
- Tsai, C.-C., & Chen, C.-C. (2006). Making decision to evaluate process capability index C_p with fuzzy numbers. *The International Journal of Advanced Manufacturing Technology*, 30(3), 334-339. doi:10.1007/s00170-005-0052-7
- Wang, C.-H. (2005). Constructing multivariate process capability indices for short-run production. *The International Journal of Advanced Manufacturing Technology*, 26(11-12), 1306-1311.
- Wang, F. K., & Chen, J. C. (1998). Capability index using principal components analysis. *Quality Engineering*, 11(1), 21-27.
- Wang, F.-K., & Du, T. C. (2000). Using principal component analysis in process performance for multivariate data. *Omega*, 28(2), 185-194.
- Westerhuis, J. A., Gurden, S. P., & Smilde, A. K. (2000). Generalized contribution plots in multivariate statistical process monitoring. *Chemometrics and Intelligent Laboratory Systems*, 51(1), 95-114.
- Xekalaki, E., & Perakis, M. (2002). *The Use of principal component analysis in the assessment of process capability indices*. Paper presented at the Proceedings of the Joint Statistical Meetings of the American Statistical Association, The Institute of Mathematical Statistics, The Canadian Statistical Society. New York.