



Universidad de Valladolid

FACULTAD DE CIENCIAS

TRABAJO FIN DE GRADO

Grado en Estadística

Estudio de Capacidad de un Proceso Industrial asociado a características cuya ley de probabilidad es normal bivalente. Aplicación al sector Aeroespacial empleando R y Minitab.

Autor: Gabriel Martín Loma

Tutor 1: Jesús Alberto Tapia García

Tutor 2: Jesús Ignacio Sesé Sánchez

Año: 2025

Agradecimientos

Agradecer a toda mi familia y a todos mis amigos por el apoyo que me han dado incondicionalmente.

Me gustaría agradecer también la participación de mis tutores que me han asesorado a la hora de realizar este trabajo.

Y agradecer a Aciturri por acogerme tan bien y por brindarme la oportunidad y las facilidades necesarias que han hecho posible este trabajo así como el conocimiento adquirido , con especial atención a Jesús Ignacio Sesé Sánchez cuya tutoría y enseñanzas me han hecho dar un gran paso para conocer el mundo laboral y al que estoy altamente agradecido.

Resumen

El objetivo de este estudio es conocer como es y cómo se realiza un proceso de mecanizado industrial concluyendo con el índice de capacidad de este. Para ello, se usaron técnicas estadísticas para describir los datos, para evaluar la normalidad de las observaciones y para detectar atípicos. Estas técnicas estadísticas son tanto numéricas con gráficas. Posteriormente se comprobará la estabilidad del proceso y finalmente se concluye con el estudio de la capacidad.

En el documento además se incluye el diagrama de flujo de las etapas de la metodología, la cual se implementa sobre un ejemplo.

Abstract

The scope of this study is to truly understand what the industrial machining process is like and how it is carried out, concluding with its capability index. To do this, statistical techniques were used to describe the data, to evaluate the normality of the observations and to detect outliers. These statistical techniques are both numerical and graphical. Additionally, the stability of the process and finally the capacity were verified.

The document also includes the flow chart of the milestones of the methodology that will be followed, since the methodology includes an example on which the study is carried out as well as the context to apply the techniques.

Índice

1	Introducción y contexto de la industria aeroespacial	7
2	Descripción del Proceso Industrial de Mecanizado	9
3	Diagrama de Flujo y Consideraciones	12
4	Técnicas Estadísticas.....	14
4.1	Proyecciones de Sammon	14
4.2	Análisis de Componentes Principales (ACP).....	15
4.3	Tests de Mardia	16
4.4	T² de Hotelling.....	17
4.5	MEWMA	18
5	Fases de la Metodología de análisis estadístico en el proceso industrial.....	19
5.1	Actuaciones Preparatorias:	19
5.2	Comprobación de supuestos.....	21
5.3	Análisis Estructural y Descriptivo de los Datos	22
5.3.1	Análisis Estructural	22
5.3.2	Análisis estadístico descriptivo	23
5.3.3	Test de Normalidad	25
5.4	Control de Calidad sobre la Muestra (Atípicos)	28
5.4.1	Distancia de Mahalanobis	28
5.4.2	Detección de atípicos mediante los Autovalores.....	29
5.5	Gráficos de Control Multivariante.....	30
5.5.1	Dispersión del Proceso	30
5.5.2	Desplazamiento de la Media.....	31
5.6	Caracterización de la Población	33
5.6.1	Estimación de los Parámetros.....	33
5.6.2	Clasificación de las Observaciones en familias	33
5.7	Índices de Capacidad del Proceso	34
5.7.1	Índice de Capacidad Tipo Ia :	36
5.7.2	Índices de Capacidad Tipos Ic y Iic	37
5.7.3	Índices de Capacidad Tipos Iia y Iib	38
6	Resultados	40
7	Conclusiones.....	52
8	Bibliografía.....	54

9	Anexos	56
9.1	Anexo A: Definiciones.....	56
9.2	Anexo B: Términos Abreviados	59
9.2.1	Acrónimos	59
9.2.2	Símbolos	60
9.3	Anexo C.....	63

1 Introducción y contexto de la industria aeroespacial

Debido a la creciente complejidad de los métodos de producción industrial, en concreto los aplicados al sector aeroespacial, y en los requisitos de calidad cada vez más exigentes para los productos resultantes de dichos procesos industriales, un análisis del proceso industrial basado en cantidades univariantes, en muchos casos, no es suficiente, aunque ésta haya sido la práctica habitual en presencia de características correlacionadas. En presencia de características correlacionadas, los resultados pueden ser engañosos y es necesario analizar el proceso teniendo en cuenta las variables conjuntamente. Esto puede ser aplicable, por ejemplo, en situaciones donde se observan tolerancias geométricas, combinaciones de presión y temperatura presentes en el conformado de materiales compuestos, o cantidades correlacionadas de materiales u otros productos (ISO 22514-6). El control de procesos en los que varias variables relacionadas son estudiadas se conoce colectivamente como control estadístico de procesos industriales multivariantes (MSPC, por sus siglas en inglés).

Este trabajo proporciona una metodología preliminar para evaluar la Capacidad de un Proceso Industrial asociado a características que siguen una distribución normal multivariante; en nuestro caso de estudio sigue una distribución normal bivariante (posición en el eje X y en el eje Y de un taladro realizado sobre una plancha metálica). En su desarrollo, se incluye los métodos de cálculo para obtener los índices de capacidad del proceso, que tienen en cuenta su dispersión y su ubicación, como una extensión de los índices correspondientes para cantidades univariantes. Estos índices propuestos se basan en los índices clásicos C_p y C_{pk} para el caso unidimensional (ISO 22514-2).

Las variaciones en el proceso industrial de fabricación con respecto al valor objetivo se recogen en dos grupos: Causas asignables (fuentes de variación conocidas, por lo tanto, acotables) y Causas no asignables (aquellas que están inherentemente en cualquier proceso industrial). Además, se asume, como condición necesaria, que el sistema de medición de los productos resultantes del proceso industrial está demostrado estar calibrado antes de determinar la capacidad del proceso a estudiar.

Los métodos que se utilizan en la industria para el control de los procesos y la evaluación de su capacidad se emplean para acercarse a procesos *Right First Time* (RFT) en los que se minimiza el número de piezas no conformes que se fabrican, así como para reducir la necesidad de inspección del 100% de las características asociadas a las piezas. Estos métodos son especialmente necesarios si: a) Se especifican valores límite para los índices de capacidad del proceso para cantidades de productos continuas y multivariadas como parte de un contrato entre clientes y proveedores. b) Se deben comparar las capacidades de diferentes instalaciones, métodos de producción o proveedores. c) Se deben aprobar procesos industriales de producción, o d) Se deben analizar problemas y tomar decisiones en casos de reclamaciones.

Este trabajo de fin de grado comienza con el diagrama de flujo como esquema general de las fases para evaluar la capacidad de un proceso industrial. Posteriormente, el documento

consta de un apartado en que se recogerán todas las técnicas estadísticas que se utilizan, un apartado en el que se contará los aspectos importantes de cada método utilizado que es necesario contar para que se entienda el porqué de su uso y un apartado de resultados. Finalmente se presentan las conclusiones. Adicionalmente, el estudio consta de los términos y simbología que se necesita para seguir el desarrollo de documento. Además, consta del esquema general del flujo para aplicar la metodología y del detalle de sus fases. Se concluye con una serie de anexos que incluyen cálculos y especificaciones, así como tablas y finalmente con la bibliografía utilizada.

Adicionalmente, este trabajo incluye la aplicación a un caso industrial que se ha llevado a cabo con lenguaje de programación R (entorno y lenguaje de programación diseñado para el análisis estadístico y la manipulación de datos), así como índices y gráficos calculados con la versión 22 de Minitab® (software de cálculo cuyo uso está muy extendido en la industria aeroespacial).

2 Descripción del Proceso Industrial de Mecanizado

En este capítulo vamos a aportar una descripción detallada del proceso de mecanizado con objeto de proporcionar una comprensión integral del mismo, destacando sus especificaciones y las observaciones obtenidas.

Suponemos un proceso industrial de mecanizado sobre una plancha metálica donde se debe realizar un taladro de 50,00 mm de diámetro con una tolerancia de $\pm 0,05$ mm, cuya posición se encuentre en un intervalo de superficie limitado por las especificaciones del plano adjunto y cuyo centro se corresponde con las posiciones: 80 mm en el eje X y -116,5 mm en el eje Y. Terminada la fabricación cada pieza producida es transportada hasta el departamento de inspección donde, tras una inspección visual satisfactoria, se mide la posición del punto central de ese taladro con un equipo apropiado. Además, se indican los límites de especificación (los límites del intervalo en los que puede variar la posición del taladro para que este sea válido):

$$L_X = 79,750 \text{ mm}$$

$$U_X = 80,250 \text{ mm}$$

$$L_Y = -116,750 \text{ mm}$$

$$U_Y = -116,250 \text{ mm}$$

Ésta es una simplificación de un verdadero proceso industrial en el que las piezas siguen un flujo de producción más complejo con múltiples operaciones de mecanizado o procesos adicionales hasta la entrega al cliente que, en la actualidad, se apoya en el empleo de la metodología APQP (Advanced Product Quality Planning - Planificación Avanzada de la Calidad del Producto), en la que se establece la gestión estandarizada en 5 fases del proyecto, considerando que el trabajo se desarrolla con un equipo multidisciplinar, de forma que se anticipan las incidencias que puedan afectar a la calidad y entrega del producto. El alcance de este proyecto se enmarca esencialmente en la Fase 4 "Validación del producto y del proceso".

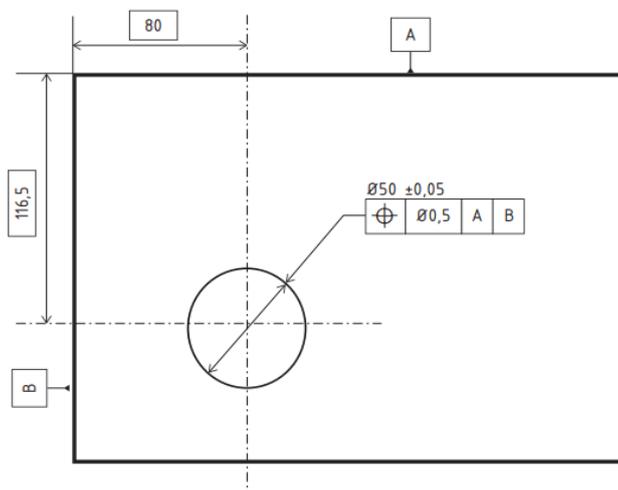


Figura 1 Representación de plano del proceso industrial

Representación gráfica del plano del proceso industrial de estudio.

El A y B son *data* que nos indican el sistema de referencia y, en nuestro caso, nos indica el sistema de coordenadas que siguen los taladros.

Después de realizar y medir las piezas a estudiar y tras un proceso de análisis y depuración de los datos disponibles se obtiene una muestra de 100 taladros cuyos pares de posición (x_i, y_i) se presentan en la Tabla 1.

Nr.	dev. D	X-coord.	Y-coord.	Nr.	dev. D	X-coord.	Y-coord.	Nr.	Dev D	X-coord.	Y-coord.
1	0,038	79,976	-116,470	36	0,090	79,995	-116,410	71	0,107	79,986	-116,394
2	0,094	79,993	-116,406	37	0,097	80,002	-116,403	72	0,073	80,016	-116,429
3	0,086	80,031	-116,420	38	0,113	80,027	-116,390	73	0,069	79,995	-116,431
4	0,041	79,968	-116,475	39	0,021	79,995	-116,520	74	0,108	79,975	-116,395
5	0,105	79,973	-116,399	40	0,085	80,010	-116,416	75	0,118	79,965	-116,387
6	0,092	79,983	-116,410	41	0,110	80,005	-116,390	76	0,122	79,971	-116,382
7	0,099	80,008	-116,401	42	0,081	80,004	-116,419	77	0,119	79,978	-116,383
8	0,086	80,014	-116,415	43	0,055	79,966	-116,457	78	0,118	79,999	-116,382
9	0,075	80,020	-116,428	44	0,097	80,013	-116,404	79	0,024	80,008	-116,477
10	0,076	79,979	-116,427	45	0,078	80,021	-116,425	80	0,094	80,005	-116,406
11	0,064	79,978	-116,440	46	0,118	79,989	-116,383	81	0,056	80,007	-116,444
12	0,086	80,016	-116,416	47	0,111	79,988	-116,390	82	0,093	80,032	-116,413
13	0,067	79,990	-116,434	48	0,057	79,987	-116,445	83	0,139	79,958	-116,368
14	0,120	79,992	-116,380	49	0,101	80,012	-116,400	84	0,122	79,990	-116,378
15	0,103	79,999	-116,397	50	0,067	80,017	-116,435	85	0,126	79,994	-116,374
16	0,119	80,016	-116,382	51	0,099	80,000	-116,401	86	0,089	80,029	-116,416
17	0,086	80,038	-116,423	52	0,101	79,995	-116,399	87	0,110	80,000	-116,390
18	0,118	80,018	-116,383	53	0,139	79,999	-116,361	88	0,084	80,010	-116,417
19	0,116	80,005	-116,384	54	0,086	80,002	-116,414	89	0,121	80,000	-116,379
20	0,118	80,071	-116,406	55	0,095	80,068	-116,433	90	0,131	79,992	-116,369
21	0,072	79,941	-116,458	56	0,103	79,990	-116,397	91	0,122	79,992	-116,378
22	0,097	79,984	-116,404	57	0,178	80,035	-116,325	92	0,062	79,990	-116,439
23	0,029	79,986	-116,475	58	0,107	79,980	-116,395	93	0,098	79,999	-116,402
24	0,093	80,043	-116,418	59	0,182	79,978	-116,319	94	0,086	79,986	-116,415
25	0,047	80,027	-116,538	60	0,099	80,000	-116,401	95	0,097	79,986	-116,404
26	0,090	80,031	-116,415	61	0,080	79,995	-116,420	96	0,092	80,020	-116,410
27	0,097	80,005	-116,403	62	0,133	79,996	-116,367	97	0,095	79,984	-116,406
28	0,122	80,024	-116,380	63	0,088	80,000	-116,412	98	0,133	79,980	-116,369
29	0,081	80,040	-116,430	64	0,107	79,948	-116,406	99	0,132	79,981	-116,369
30	0,094	80,006	-116,406	65	0,101	80,015	-116,400	100	0,058	80,033	-116,452
31	0,099	79,986	-116,402	66	0,081	79,990	-116,420				
32	0,094	79,982	-116,408	67	0,087	80,009	-116,413				
33	0,111	79,942	-116,405	68	0,067	80,004	-116,433				
34	0,135	79,975	-116,367	69	0,130	79,960	-116,376				
35	0,103	80,014	-116,398	70	0,121	80,007	-116,379				

Tabla 1 Observaciones

Las columnas de la Tabla 1 son el número de observación, las desviaciones (distancia en valor absoluto) con respecto del valor nominal (donde deberían estar realmente las observaciones), y las posiciones en los ejes X e Y. Desviaciones y posiciones están expresadas en mm. Estos serán los datos con los que se trabajará en los puntos siguientes y de los que se va a estudiar la capacidad para saber cómo es el proceso.

3 Diagrama de Flujo y Consideraciones

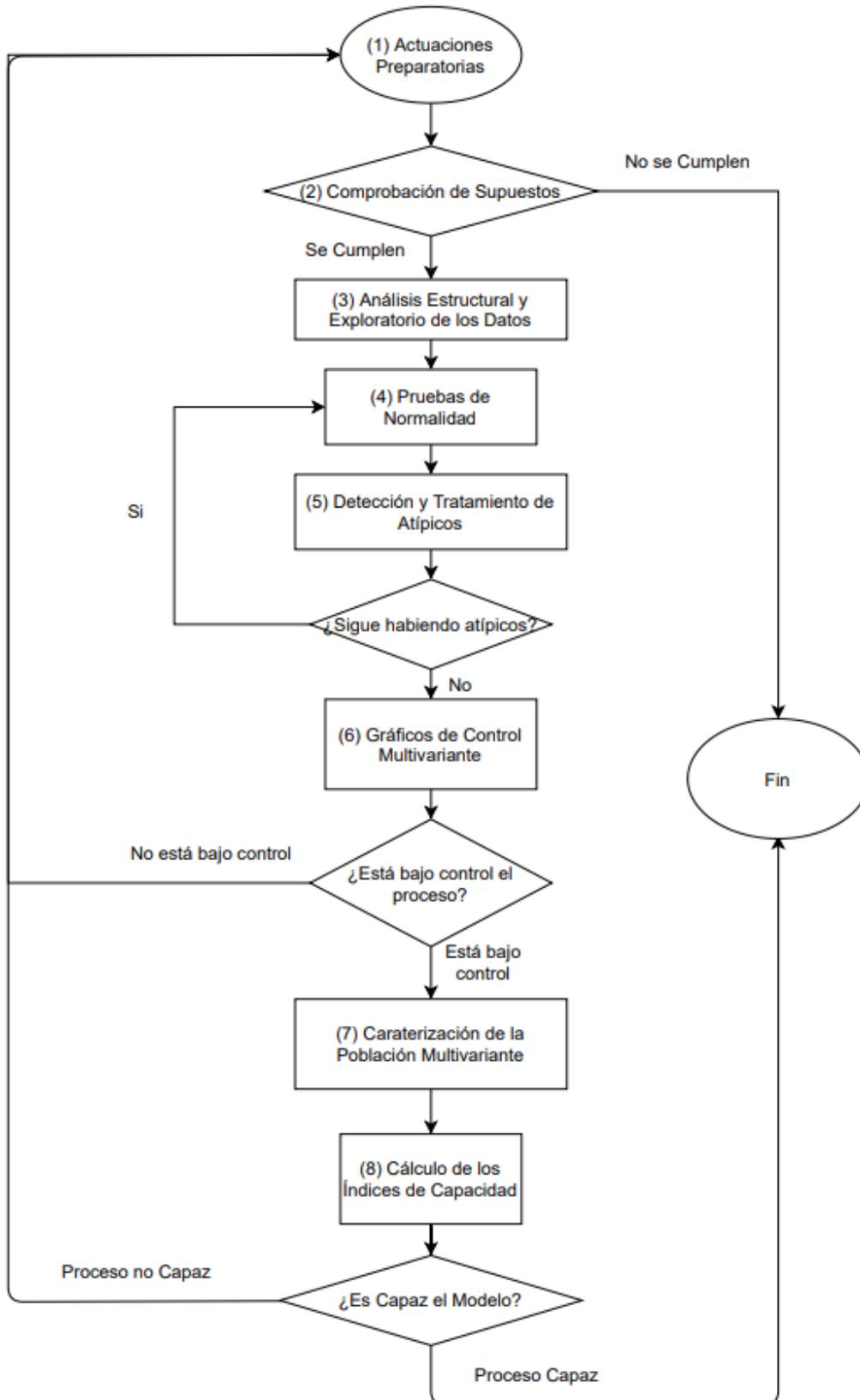


Figura 2 Diagrama de Flujo de la metodología propuesta

El diagrama de flujo presentado en la Figura 2 muestra la metodología empleada para evaluar la capacidad de un proceso industrial multivariante.

Las fases que se seguirán en la metodología propuesta para comprobar la capacidad del proceso son las siguientes: Tras caracterizar y comprender los condicionantes del proceso industrial, para comprobar la capacidad del proceso y haber establecido los elementos de naturaleza estadística apropiados para el estudio [recogido en el punto (1) Actuaciones preparatorias] se procede a la comprobación de los supuestos requeridos para esta metodología [(2)]. Una vez que el proceso industrial de estudio está correctamente definido, se continúa entendiendo como es nuestra muestra [(3) Análisis Estructural y Exploratorio de los Datos], que incluye también la evaluación de la normalidad [(4) Pruebas de normalidad]. Posteriormente llega la fase de control de calidad sobre la muestra [(5) Detección y tratamiento de atípicos] y, se continúa evaluando la estabilidad del proceso empleando gráficos de control [(6) Gráficos de control multivariante]. Una vez que se tiene el proceso de mecanizado refinado, es necesario conocer la población de origen de tus datos [(7) Caracterización de la población multivariante] para poder concluir con la última fase [(8) Cálculo de los índices de capacidad].

4 Técnicas Estadísticas

En este capítulo se introducen formalmente las técnicas que se necesitan para obtener la información que se busca.

4.1 Proyecciones de Sammon

El método de Sammon obtiene la distribución de las proyecciones que proporcione el valor mínimo optimizado para las discrepancias E .

$$E = \frac{1}{\sum_{i=1}^{n-1} \sum_{j=i+1}^n \delta_{ij}} \sum_{i=1}^{n-1} \sum_{j=i+1}^n \frac{(\delta_{ij} - d_{ij})^2}{\delta_{ij}}$$

Siendo δ_{ij} la distancia entre las observaciones \vec{x}_i y \vec{x}_j , y d_{ij} la distancia entre sus proyecciones

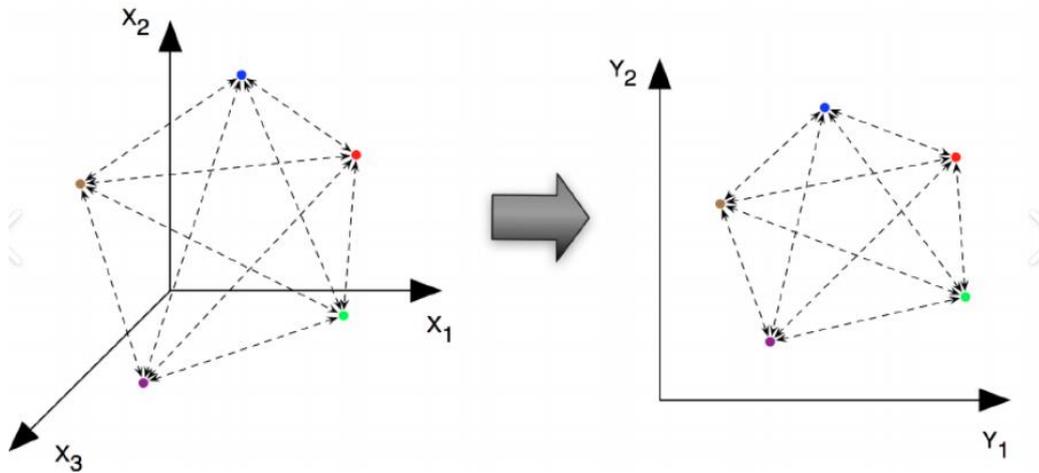


Figura 3 Ejemplo proyecciones de Sammon

Uno de los posibles métodos para encontrar el valor óptimo es el denominado *steepest descent algorithm*; sin embargo, tiene un coste computacional ($O(N^2)$) creciente con el cuadrado del tamaño de la muestra N ; además, si se desea ubicar una nueva observación debe recalcularse de nuevo el algoritmo. En cada iteración la solución se mueve en la dirección del gradiente de la función de discrepancia E .

$$X_i^{m+1} = x_i^m - c \frac{\left| \frac{\partial E^{(m)}}{\partial x} \right|_{x=x_i^{(m)}}}{\left| \frac{\partial^2 E^{(m)}}{\partial x^2} \right|_{x=x_i^{(m)}}}$$

$$y_i^{m+1} = y_i^m - c \frac{\left| \frac{\partial E^{(m)}}{\partial y} \right|_{y=y_i^{(m)}}}{\left| \frac{\partial^2 E^{(m)}}{\partial y^2} \right|_{y=y_i^{(m)}}}$$

Donde c es una constante que modifica la velocidad de convergencia del método y $proyección_{2D}(\bar{x}_i) = (x_i, y_i)$.

4.2 Análisis de Componentes Principales (ACP)

Se trata de buscar un sistema ortogonal de nuevas coordenadas expresadas como $w = \{\bar{w}_1, \bar{w}_2, \dots, \bar{w}_N\}$ que se denominan componentes principales. Sea la matriz de cambio de base $L = \{l_{ij}\}$. Los coeficientes que nos relacionan la base antigua con la nueva son:

$$\bar{w}_i = \sum_{j=1}^n l_{ij} \bar{x}_j, \quad i = 1, \dots, n$$

La condición de ortogonalidad (no-correlación) nos lleva a

$$E(\bar{w}_i \cdot \bar{w}_j) = E\left(\sum_{k=1}^n l_{ik} \bar{x}_k \sum_{m=1}^n l_{jm} \bar{x}_m\right) = 0, \quad (i \neq j)$$

O, lo que es lo mismo,

$$\sum_{k,m=1}^n l_{ik} l_{jm} E(\bar{x}_k \bar{x}_m) = \sum_{k,m=1}^n l_{ik} l_{jm} c_{km} = 0$$

La nueva base debe ser ortogonal, lo que se consigue imponiendo

$$\sum_{i=1}^n l_{ij} l_{ik} = \delta_{ik}$$

Siendo δ_{ik} la delta de Kronecker.

Con estas condiciones el problema está suficientemente condicionado. Para concretar aún más su solución se puede escribir de forma matricial:

$$E[\bar{w}\bar{w}^T] = E[L\bar{x}(L\bar{x})^T] = E[L\bar{x}\bar{x}^T L^T] = LcL^T = \Lambda$$

Donde Λ es la matriz de varianzas-covarianzas en la nueva base, y es diagonal, por lo que el problema se reduce a diagonalizar la matriz de varianzas-covarianzas Σ .

$$|\Sigma - \lambda I| = 0$$

Donde los autovectores son las nuevas componentes principales y sus autovalores asociados son el “peso” o “importancia”.

Existen varios criterios de selección de r , con $r < n$, componentes principales:

1. Criterio de % de pérdida de variación:

La suma de las desviaciones al cuadrado de la matriz de covarianzas inicial C y la estimada basada en el subconjunto de r componentes viene dada por

$$tr(C^t C) - \sum_{j=1}^r \lambda_j = \sum_{j=r+1}^n \lambda_j$$

de ahí que la proporción de la suma total de desviaciones al cuadrado debida a las primeras r componentes, ip (información preservada) para determinar el número de componentes que a preservar se establece un valor y que se expresa como:

$$ip = 100 * \frac{\sum_{j=1}^r \lambda_j}{\sum_{j=1}^n \lambda_j}$$

2. Criterio de la media aritmética:

Dado que la variación total viene dada por la suma de los λ_j , donde λ_j es la varianza de w_j , un posible criterio es retener aquellas componentes cuya varianza supere la media aritmética.

3. Criterio de la media geométrica:

Este criterio selecciona los componentes a preservar cuyos autovalores sean superiores a la media geométrica de los mismos.

4.3 Tests de Mardia

Si se definen los coeficientes multivariados de asimetría y kurtosis.

$$b_{1,n} = \frac{1}{N^2} \sum_{i=1}^N \sum_{j=1}^N m_{ij}^3$$

$$b_{2,n} = \frac{1}{N} \sum_{i=1}^N m_i^2$$

Donde

$$m_i = (x_i - \bar{x})^t S^{-1} (x_i - \bar{x})$$

$$m_{ij} = (x_i - \bar{x})^t S^{-1} (x_j - \bar{x})$$

4.4 T^2 de Hotelling

Para este apartado y para el apartado 5.5 es importante empezar definiendo las dos fases en las que se divide la producción.

- I. Fase I: En esta fase se comprueba que los datos obtenidos están bajo lo que se entiende como un proceso industrial estadísticamente bajo control.
- II. Fase II: En esta fase se comprueba que el proceso sigue bajo control para las observaciones futuras.

El gráfico de control se basa en la secuencia del siguiente estadístico:

$$T_j^2 = (X_j - \bar{x})^T S^{-1} (X_j - \bar{x}), \text{ donde } j = 1, 2, \dots, m.$$

Entonces el estadístico $T_j^2 / d_0(m)$ sigue una distribución beta con $d/2$ y $\frac{1}{2} \left(\frac{2(m-1)^2}{3m-4} - d - 1 \right)$ grados de libertad, donde $d_0(m) = (m-1)^2 m^{-1}$.

El límite superior para el gráfico de control multivariante de Shewhart es:

$$U_{CL} = \frac{(m-1)^2}{m} B_{1-\alpha, d/2, \left(\frac{2(m-1)^2}{3m-4} - d - 1 \right) / 2}$$

Para la fase II, siendo x_f una observación futura independiente de \bar{x} y de S , el gráfico de control se basa en el siguiente estadístico:

$$T_j^2 = (X_f - \bar{x})^T S^{-1} (X_f - \bar{x})$$

El estadístico $T_f^2 / d_1(m, d)$ sigue una distribución F con d y $(m-d)$ grados de libertad, donde $d_1(m, d) = d(m+1)(m-1)[m(m-d)]^{-1}$ donde m es el número de observaciones utilizadas en la fase I.

Por lo tanto, el gráfico de control multivariante de Shewhart tiene un límite de control superior (U_{CL}):

$$U_{CL} = \frac{d(m+1)(m-1)}{m(m-d)} F_{1-\alpha, d, m-d}$$

Se debe considerar también, aunque no es el caso de nuestros datos, el supuesto en el que el tamaño de los subgrupos racionales sea mayor que 1 ($n > 1$), para la fase I se calculan $\bar{\bar{x}}$ y $\bar{\bar{S}}$ como la media de las medias de los subgrupos y la matriz de varianzas-covarianzas de las medias de los subgrupos. En este supuesto el estadístico nos queda cómo:

$$T_j^2 = n(\bar{x}_j - \bar{\bar{x}})^T \bar{\bar{S}}^{-1}(\bar{x}_j - \bar{\bar{x}}), \text{ donde } j = 1, 2, \dots, m.$$

Para el subgrupo j -ésimo. El estadístico $T_j^2 / c_0(d, m, n)$ sigue una distribución F con d y $(mn - m - d + 1)$ grados de libertad puesto que $c_0(d, m, n) = d(m - 1)(n - 1)(mn - m - d + 1)^{-1}$

Y el límite de control superior es:

$$U_{CL} = \frac{d(m - 1)(n - 1)}{mn - m - d + 1} F_{1-\alpha, d, mn-m-d+1}$$

Para la fase II,

$$T_f^2 = n(\bar{x}_f - \bar{\bar{x}})^T \bar{\bar{S}}^{-1}(\bar{x}_f - \bar{\bar{x}})$$

Donde el estadístico $T_f^2 / c_1(d, m, n)$ sigue una distribución F con d y $(mn - m - d + 1)$ grados de libertad puesto que $c_1(d, m, n) = d(m - 1)(n - 1)(mn - m - d + 1)^{-1}$

Y el límite de control superior es:

$$U_{CL} = \frac{d(m - 1)(n - 1)}{mn - m - d + 1} F_{1-\alpha, d, mn-m-d+1}$$

4.5 MEWMA

El estadístico MEWMA es la extensión lógica del EWMA en el caso multivariante y se define como:

$$Z_j = \lambda x_j + (1 - \lambda)Z_{j-1} \quad \text{con } j = 1, 2 \dots$$

Donde $Z_0 = \mu_0$, la media de bajo control y,

$$\lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_d) \quad 0 < \lambda_i \leq 1$$

Generalmente se asume que todos los λ_i son iguales y que $\lambda_i = \lambda$ por lo que el estadístico MEWMA se puede escribir como:

$$Z_j = \lambda x_j + (1 - \lambda)Z_{j-1} \quad \text{con } j = 1, 2 \dots$$

La constante λ es el parámetro móvil del EWMA y este se reduce a x_j cuando $\lambda = 1$.

El gráfico de control MEWMA nos da señal de fuera de control cuando:

$$Y_j^2 = (Z_j - \mu_0)^T \Sigma_{Z_j}^{-1} (Z_j - \mu_0) > h$$

Donde h representa el límite de control superior y Σ_{Z_j} la matriz de varianzas-covarianzas del estadístico Z_j es:

$$\Sigma_{Z_j} = \frac{\lambda}{2-\lambda} [1 - (1-\lambda)^{2j}] \Sigma_0$$

Este resultado es análogo al caso univariante EWMA. Cuando $\lambda = 1$, el gráfico se reduce a un gráfico de control tipo Shewhart.

5 Fases de la Metodología de análisis estadístico en el proceso industrial

5.1 Actuaciones Preparatorias:

En esta fase inicial ($\{1\}$) se trata de definir y entender las condiciones del estudio a realizar, esto incluye desde determinar cuáles son las características de interés de la pieza, las unidades del sistema de medida, el número de muestras del que se dispone, los detalles tanto del proceso de fabricación como del proceso de medición, determinar el nivel de significación a utilizar en los test (habitualmente con un nivel de confianza del 95% aunque en algún caso puede variar en función de las normas internacionales, requerimientos del cliente o, inclusive, en caso de necesidad bajo la justificación correspondiente), etc. También es apropiado señalar la naturaleza de las características del estudio (según la norma internacional ISO 1101 los diferentes tipos son: Forma, Orientación, Localización y Alabeo) y sus tolerancias con especial atención a modificadores singulares que se incorporan en los planos para facilitar el ensamblaje posterior de la pieza. Estos modificadores aplicados sobre los límites de especificación, hacen variables estos límites; sirva como, por ejemplo, “la condición de máximo material” representada en los planos con el símbolo \textcircled{M} .

Además, es apropiado indicar el tamaño de los subgrupos racionales (n), el tipo de gráficos de control que se van a utilizar y el tipo de índices a seguir para el cálculo de la capacidad que se desarrollarán más adelante.

El caso de estudio es el siguiente:

$$\text{Matriz de los datos } X = (x_{ij})_{i=1, \dots, N}^{j=1, \dots, d} = \begin{pmatrix} x_{11} & \cdots & x_{1d} \\ \vdots & \ddots & \vdots \\ x_{N1} & \cdots & x_{Nd} \end{pmatrix}$$

Donde $i = 1, \dots, 100$ representa la observación i -ésima de las $N = 100$ observaciones, de la tabla anterior. La dimensión de nuestros datos $d = 2$ y nuestro caso no consta de subgrupos racionales puesto que no tenemos ninguna variable factor, por lo que el tamaño del subgrupo $n = 1$.

Es momento también de determinar los gráficos de control apropiados para el caso, habitualmente en el caso multivariante son T^2 de Hotelling, varianza generalizada y MEWMA

(aunque el desarrollo de estos gráficos de control se comentará con detalle en su fase correspondiente, en esta fase únicamente se trata de mencionar cuales son los apropiados).

De los diferentes tipos de índices que se desarrollarán en el apartado 5.5 se elige el tipo *Ia* que consiste en que se transforma la región de tolerancia a la región de variación del proceso industrial y se calcula el índice mediante probabilidades (se explicarán los diferentes métodos de cálculo más adelante, en el apartado 5.5).

5.2 Comprobación de supuestos

El objetivo de la fase de comprobación de los supuestos {2} es asegurar que se cumplen los requisitos para poder aplicar los métodos estadísticos correspondientes porque de lo contrario los resultados no podrían ser interpretables. Teniendo en cuenta lo dicho, ésta es una fase clave pues, si no es posible asegurar los supuestos (como la independencia de las observaciones o la correlación entre las variables), es posible que no se pueda seguir adelante con el análisis o que se tengan que aplicar otros métodos diferentes apropiados a ese caso de estudio.

No por evidente debe omitirse el primer supuesto a comprobar: las características de estudio deben ser variables continuas y, por tanto, los datos sean datos numéricos continuos puesto que, si nos encontramos con datos categóricos no se podría llevar a cabo los métodos descritos en este estudio.

El segundo supuesto a revisar es la independencia de las observaciones (aunque en la mayoría de los casos se acepte por la propia naturaleza del proceso de fabricación).

El tercer supuesto a comprobar requiere verificar la existencia de una correlación fuerte entre las variables. Típicamente se considerará como una correlación alta y por lo tanto válida para seguir con el estudio cuando $\rho \geq |0,7|$; sin embargo, en algunos casos quizás es necesario realizar test para evaluar la correlación ($H_0: \rho = 0, H_1: \rho \neq 0$).

El cuarto supuesto exige comprobar la normalidad de las observaciones pues esto permite que algunos métodos que se describirán más adelante puedan ser efectuados.

El quinto y último supuesto a confirmar, obligatorio para evaluar el control del proceso, es la disponibilidad de los datos en el orden cronológico de fabricación de la característica que se estudia. Este supuesto, que parece una obviedad, debe confirmarse pues en el caso de que una pieza salga del flujo de fabricación estándar, en un momento posterior a la fabricación de la característica de interés, por requerir algún retrabajo o tratamiento especial sobre una zona diferente a la que se estudia en el proceso industrial, es habitual que pase por la fase de inspección fuera de la posición que le correspondería en el flujo de producción en el que se produjo la característica, por lo que en varias ocasiones no coincide el orden de fabricación con el orden de inspección; algo tan simple como ordenar los datos del estudio por fecha de fabricación, en vez de por fecha de inspección evita sorpresas indeseadas..

5.3 Análisis Estructural y Descriptivo de los Datos

En esta etapa 3 se pretende extraer la información que contienen los datos, describiendo las variables y empleando, cuando sea posible, técnicas gráficas con objeto de mostrar si existe dependencia entre las variables y así poder analizar la normalidad de los datos con técnica de bondad de ajuste.

5.3.1 Análisis Estructural

Para continuar con el análisis estructural de las observaciones emplearemos dos métodos: Las proyecciones de Sammon y ACP.

5.3.1.1 Proyecciones de Sammon

Existen distintos algoritmos matemáticos que permiten examinar, de forma visual clara, la disposición espacial de los N puntos correspondientes a espacios vectoriales d -dimensional, a través de su proyección sobre un espacio vectorial de dimensión 2 (aunque nuestro caso al ser $d = 2$, se harán referencias a $d > 2$ por su interés en otros casos). De estos algoritmos se muestran como más útiles aquellos que intentan conservar las distancias entre los distintos puntos a la hora de proyectar, de forma que la distancia entre dos puntos cualesquiera pueda ser medible, si así se desea, directamente sobre su proyección (Manuel Castejón Limas, Joaquín Ordieres Meré, Francisco Javier de Cos Juez, Francisco Javier Martínez de Pisón, 2014).

La capacidad de ver sobre un plano la posición de cada punto permite distinguir visualmente las distintas clases de comportamiento que pudieran presentarse, por aparecer agrupados los puntos en la proyección, así como también desestimar (o estudiar separadamente) vectores demasiado alejados a cualquier grupo reconocible. Las diferentes clases estarán formadas por aquellas observaciones que presenten un comportamiento uniforme entre sí y distinto al de otros grupos existentes.

Hay varios métodos para examinar la disposición espacial de los datos a través de sus proyecciones. El método de Sammon obtiene la distribución de las proyecciones que proporcione el valor mínimo optimizado para las discrepancias E .

Sin embargo, el método es computacionalmente costoso por lo que algunas soluciones para aumentar la velocidad del algoritmo, cuya idea básica radica en coger un subconjunto de puntos y calcular sobre ellos el algoritmo, añadiendo el resto de puntos con posterioridad, son:

1. Triangulación: Este método se fundamenta en el hecho de que cuando un nuevo punto es proyectado, sus distancias a dos puntos previamente proyectados sobre el plano se conservan.
2. Redes Neuronales: El patrón de entrenamiento considera como variables de entrada las coordenadas de los puntos en el espacio d -dimensional, y como salida las coordenadas (x, y) de sus proyecciones. Tras el entrenamiento de la correspondiente red se puede aproximar la ubicación de nuevos puntos por

los valores que la red estime en sus salidas tras presentar en la entrada las coordenadas de estos nuevos puntos.

3. Transformación Algebraica: Otra posible solución radica en una transformación lineal de la matriz de distancias en el espacio d-dimensional de forma que su transformada sea la matriz de distancias de las proyecciones obtenidas con el Sammon. Tras obtener el operador se puede aplicar a nuevos puntos de forma inmediata.

No hace falta hablar de proyecciones para dos dimensiones puesto que se pueden observar directamente las observaciones en todas sus dimensiones, para nuestros datos nos sirve con representarlos.

5.3.1.2 Análisis de Componentes Principales (ACP)

Para conocer cómo varían nuestros datos e identificar el menor número posible de variables que los pueden caracterizar nuestra muestra existe la técnica de ACP, que fundamentalmente nos indica las direcciones de máxima variabilidad dentro de nuestros datos y que nos ayudará a determinar cuáles son las variables (combinación lineal de variables originales) que más aportan cuantitativamente a la variabilidad del proceso (Manuel Castejón Limas, Joaquín Ordieres Meré, Francisco Javier de Cos Juez, Francisco Javier Martínez de Pisón, 2014).

El análisis fundamental proporciona las direcciones de la base óptima de acuerdo con criterios de mínimos cuadrados. Frecuentemente se suele utilizar como técnica de reducción dimensional puesto que nos indica las orientaciones preferentes y desprecia aquellas que influyen menos en la variable a explicar.

La diferencia entre usar la matriz de covarianzas o de correlaciones viene dada por la posibilidad de comparar las variables porque tienen la misma unidad y escala. Esta diferencia en el ACP se denomina ACP normado (matriz de correlación) y ACP no normado (matriz de covarianzas). Se utiliza el método ACP para conocer más acerca los datos como puede ser las direcciones máxima variabilidad, no solo para saber la información que aportan las dimensiones principales.

5.3.2 Análisis estadístico descriptivo

Una vez que tenemos bien identificados nuestros datos (Tabla 1) y se conoce su estructura se continua con el cálculo de los estadísticos muestrales y con el análisis de la distribución de cada componente (con especial atención a la distribución normal, dadas la exigencia del supuesto de normalidad en las herramientas estadísticas.

5.3.2.1 Análisis descriptivo numérico

Para describir los datos de la Tabla 1, se comienza calculando el vector de medias, la matriz de varianzas-covarianzas y la matriz de correlaciones. Adicionalmente, para generar los gráficos de control nos encontraremos con la necesidad de trabajar con otros estadísticos como el T^2 de Hotelling, etc. Estos estadísticos se calcularán a medida que sean necesarios en el desarrollo del estudio.

La matriz de covarianzas Σ de la población proporciona conocimiento acerca de la relación existente entre las distintas variables. Cuando la matriz de covarianzas es diagonal nos indica que las variables son independientes entre sí, siempre que se cumplan condiciones de normalidad multivariante. Si además los elementos de la diagonal fueran iguales unos a otros, $\Sigma = \sigma \cdot I$, se estaría frente a una distribución hiper-esférica de datos.

Se calculan los siguientes estadísticos muestrales: vector de medias \bar{x} , matriz de varianzas-covarianzas Σ y la matriz de correlaciones ρ .

5.3.2.2 Análisis descriptivo grafico

Los gráficos comúnmente denominados como *box-plot* son una herramienta útil a la hora de resumir la información de forma compacta donde, la caja rectangular representa el rango intercuartílico $Q1 - Q3$ con una línea situada en la mediana de los datos, con su determinado intervalo de confianza (IC). Las líneas exteriores se extienden hasta 1.5 veces de la longitud del rango intercuartílico, donde su final se indica en la posición del mínimo o del máximo, sin embargo, si el mínimo/máximo se encuentra fuera de este rango (potenciales atípicos) se representan con un asterisco (Manuel Castejón Limas, Joaquín Ordieres Meré, Francisco Javier de Cos Juez, Francisco Javier Martínez de Pisón, 2014).

Las gráficas de frecuencias totales y por clases representan la distribución que sigue cada variable de la muestra completa considerando de forma conjunta todas las clases o para cada clase.

5.3.3 Test de Normalidad

La normalidad del comportamiento de la distribución es de importancia dado el gran número de herramientas que requieren de la condición de normalidad multivariantes para su validez. Por eso, en el diagrama de flujo, se encuentra en una fase diferenciada {4} (Manuel Castejón Limas, Joaquín Ordieres Meré, Francisco Javier de Cos Juez, Francisco Javier Martínez de Pisón, 2014).

La normalidad se puede evaluar empleando el test de normalidad bivalente directos (como son los Tests de Mardia), adicionalmente, puesto que la normalidad multivariante se cumple si cada una de las d componentes que forman la combinación lineal siguen una distribución normal univariante (porque si Z es un vector cuyas componentes z_i siguen cada una una distribución normal $N(\mu, \sigma)$ y A es una matriz de coeficientes constantes que conforma la combinación entonces la X con $X = A * Z$ sigue una distribución normal multivariante), se comprobará con test de normalidad de KS en cada una de sus componentes.

Dentro de las técnicas para comprobar la normalidad existen tanto técnicas gráficas como estadísticas.

5.3.3.1 Técnicas Gráficas

Las curvas de probabilidad acumulada y la recta cuantil-cuantil son los tipos de gráficos más utilizados para la comparación de dos muestras (Stephans, M.A. 1974).

El primero consiste en trazar la distribución acumulada de tus datos y la distribución acumulada de una distribución normal y así comparar gráficamente las similitudes que comparten.

Para representar los cuantiles observados frente a los esperados se utiliza la recta cuantil-cuantil. Dada una muestra de N observaciones $\bar{z}_1, \bar{z}_2, \dots, \bar{z}_N$ de una variable aleatoria unidimensional, se denominan cuantiles a los valores de estas observaciones reordenados de menor a mayor valor $\bar{z}_{(1)}, \bar{z}_{(2)}, \dots, \bar{z}_{(N)}$.

Representando gráficamente los valores observados frente a los esperados debería obtenerse una recta del tipo identidad ($y=x$), por lo que si la recta no corresponde a la identidad (o si los datos no conforman una recta), los datos no se ajustan a la distribución esperada (en nuestro caso la normal).

Algunas anomalías que puede presentar esta recta son: Si no pasa por el origen significa que la distribución está desplazada, si la pendiente no es unitaria es que existe un factor de escala, un trazo anormalmente brusco en los extremos indica la presencia de atípicos, curvatura en la recta nos indica kurtosis y concavidad o convexidad nos indica la falta de simetría en la distribución de la muestra.

5.3.3.2 Análisis estadístico de la normalidad

Para entender los test de normalidad hay que primero saber que son los test de contraste de hipótesis (Ralph B. D'Agostino, Michael A. Stephens 1997).

En el caso de las técnicas estadísticas de contraste de hipótesis, éstas se rechazarán o no en función del valor que adopten los estadísticos seleccionados, y del grado de confianza que deseemos adoptar en nuestras decisiones.

Existen los test de contraste paramétricos y no paramétricos. Para poder aplicar los contrastes paramétricos es necesario que los datos sigan una distribución normal, mientras que los contrastes no paramétricos serán válidos, aunque la distribución de los datos se aleje de la normal.

5.3.3.2.1 Tests de Mardia

Los Tests de Mardia nos permiten calcular cuánto se ajusta la distribución normal a nuestros datos mediante los estadísticos observados de asimetría y kurtosis y sus respectivos p-valores (Ralph B. D'Agostino, Michael A. Stephens 1997) (Jobson 1998).

Si se definen los coeficientes multivariados de asimetría y kurtosis, $b_{1,n}$ y $b_{2,n}$ respectivamente.

Entonces, el coeficiente $\frac{N}{6}b_{1,n} \sim \chi^2$ con $\frac{n(n+1)(n+2)}{6}$ grados de libertad y, también $b_{2,n} \sim N(n(n+2), \frac{8n(n+2)}{N})$, sin embargo, el segundo coeficiente es más robusto para muestras grandes. Ambos coeficientes son invariantes por transformaciones lineales y se utilizarán como estadísticos para el test de normalidad.

Sin embargo, por el elevado coste computacional de los métodos anteriores, habitualmente se optará por otros test no paramétricos como son χ^2 , Saphiro-Wilk, Kolmogorov-Smirnov, Anderson-Darling y Ryan-Joiner, en este estudio sólo se llevará a cabo el Kolmogorov-Smirnov, comprobando la normalidad de cada componente.

5.3.3.2.2 Análisis de la normalidad de las componentes principales

Puede demostrarse que, si las observaciones correspondientes a cada una de las componentes principales siguen una distribución normal univariante, el conjunto sigue una distribución normal multivariante.

El test de Kolmogorov-Smirnov se basa en la comparación de la función de probabilidad observada y la teórica. Primero se calcula el estadístico normalizado $z_i = \frac{x_i - \bar{x}}{s}$ y se obtiene la función de probabilidad acumulada considerando que sigue una distribución normal.

El estadístico D multiplicado por el factor $(\sqrt{N} - 0.01 + \frac{0.85}{\sqrt{N}})$ usando como valores críticos que se propuso Stephens, M.A. 1974 en EDF Statistics for Goodness of fit. Siendo $D = \max \left\{ \left| \frac{i}{N - F(i)} \right| \right\}$, $i = 1, 2, \dots, N$.

Aplicando las desigualdades de Bonferroni (Si H_1, H_2, \dots, H_N representan las afirmaciones " \vec{w}_1 es normal", " \vec{w}_2 es normal", ..., " \vec{w}_N es normal", donde \vec{w}_i son las componentes principales) se verifica.

$$p(H_1 \cap \dots \cap H_N) \geq 1 - \sum_{i=1}^N p(\bar{H}_i)$$

Siendo $p(\bar{H}_i)$ la probabilidad de que se rechace \bar{H}_i cuando es cierta y siendo $p(H_1 \cap \dots \cap H_N)$ la probabilidad de que las afirmaciones sean aceptadas cuando son ciertas. Si ponemos que el nivel de significación sobre cada componente sea $\frac{\varepsilon}{N}$ entonces la probabilidad de acertar al afirmar que la distribución es normal es:

$p(H_1 \cap \dots \cap H_N) \geq 1 - N \frac{\varepsilon}{N} = 1 - \varepsilon$, lo que se traduce en que el nivel de significación para esta prueba conjunta es menor que ε .

5.4 Control de Calidad sobre la Muestra (Atípicos)

Los espúreos son observaciones cuya disposición espacial resulta considerablemente distinta frente al comportamiento general. Esta inconsistencia puede deberse a varios factores como la contaminación de la muestra o errores en el muestreo, valores extremos de la distribución original o provocados por la existencia de un nuevo grupo de comportamiento diferenciado (Manuel Castejón Limas, Joaquín Ordieres Meré, Francisco Javier de Cos Juez, Francisco Javier Martínez de Pisón, 2014).

Este análisis {5} para detectar atípicos debe ser realizado dentro de cada clase puesto que las diferencias entre grupos suelen ser apreciables y un punto típico de una clase, con toda probabilidad puede tener un comportamiento claramente anómalo frente a otra clase distinta a la suya, sin embargo, en el caso objeto de estudio no se realiza por la falta de existencia de clases.

En ocasiones se interpretará este apartado cómo un proceso iterativo puesto que si, tras el tratamiento de atípicos, se comprueba que siguen quedando se volverá a realizar esta fase, una vez comprobado nuevamente la bondad de ajuste a la distribución normal de los datos.

En este estudio se plantean dos métodos para el cálculo de observaciones atípicas: 1) Mediante distancias de Mahalanobis y 2) Mediante ACP.

5.4.1 Distancia de Mahalanobis

Definimos las distancias generalizadas de Mahalanobis como

$$d_j^2 = (x_j - \bar{x})S^{-1}(x_j - \bar{x})^t \text{ donde } \bar{x} \text{ es el centro de la muestra.}$$

Los N valores de las distancias d_j^2 siguen una distribución χ^2 con n grados de libertad. Si una muestra concreta se caracteriza por encontrarse a una distancia superior a la dada por el percentil determinado de la distribución χ^2 para un determinado nivel de confianza, se acepta que esta observación se trata de un *outlier*.

Este método no detecta únicamente aquellos puntos que por su localización especial serían detectados por las proyecciones de Sammon, sino que aquellos que aun siendo espúreos, quedan confundidos dentro de la nube de puntos.

Se representan los valores d_j^2 en la Figura 11.

5.4.2 Detección de datos atípicos mediante los Autovalores

En el análisis de componentes principales se determinan aquellas variables que influyen en mayor medida al comportamiento de las variables, siendo estas variables incorreladas. La interpretación del ACP para detectar espúreos viene dada por la cantidad de información que aportan las $(p - r)$ componentes que contribuyen con menor variabilidad. En el caso que de dicha información sea significativa, aunque se encuentre en el medio de la nube de puntos, se tratará de una observación atípica.

Para la observación j -ésima expresada en sus componentes principales x_{jk} , la suma de cuadrados de sus componentes principales divididas entre su varianza, $\sum_{k=1}^n \frac{x_{jk}^2}{\lambda_k}$, coincide con la distancia de Mahalanobis de la observación al centro de la muestra. Entonces, si se considera la suma parcial de las últimas $(p - r)$ componentes, $\sum_{k=r+1}^n \frac{x_{jk}^2}{\lambda_k}$, se puede determinar cuantitativamente la correspondiente variación que aportan estas últimas componentes en la observación y, si la variación se distribuye preferentemente en estas últimas componentes significa que esta observación se tratará de un atípico con respecto a la estructura de correlación.

El estadístico $F_j^2 = \sum_{k=r+1}^n \frac{x_{jk}^2}{\lambda_k}$ de los autovalores no considerados y representando, aquellos valores que queden por encima del nivel de significación marcado por el valor $\chi_{n-r}^2(\alpha)$ son observaciones espúreas.

Nuevamente, en nuestro caso no tiene sentido realizar esta prueba dado que ambas componentes principales tienen un peso elevado en términos de variabilidad, pero se indica el método para otros problemas en los que se pretenda aplicar esta metodología de actuación.

5.5 Gráficos de Control Multivariante

Las herramientas más útiles para el control estadístico de procesos multivariados son los gráficos de control multivariantes (ISO 13003). Estos gráficos nos permiten evaluar y controlar procesos estadísticos teniendo en cuenta la dependencia entre las características de calidad (ISO 7870-7).

La función de un sistema de control estadístico de procesos es proporcionar una señal estadística cuando hay causas asignables de variación presentes. La eliminación sistemática de estas causas, a través de esfuerzos continuos, lleva al proceso industrial a un estado de control estadístico. Una vez que el proceso industrial opera en control estadístico, su desempeño es predecible y su capacidad para cumplir con las especificaciones puede evaluarse (ISO 3534-2).

Los gráficos de control multivariantes funcionan mejor cuanto menor sea el número de variables, generalmente 10 o menos, por lo que a medida que aumenta el número de variables, los gráficos de control pierden eficiencia en la detección de cambios en el proceso de mecanizado. En estas situaciones, un enfoque popular es reducir la dimensionalidad de los datos con técnicas mencionadas anteriormente como ACP. Aunque para realizar este enfoque es necesario poseer un conjunto de datos históricos que se asumen que están bajo control.

Esta fase se trata de una fase crítica puesto que en el caso de que su resultado sea negativo, se requiere evaluar las causas asignables que han causado que el proceso industrial esté fuera de control.

Los gráficos de control se componen de:

1. Gráfico de control para el desplazamiento de la media. En los que se incluyen T^2 de Hotelling y MEWMA.
2. Gráfico de control multivariante para la dispersión del proceso, como el gráfico de varianzas generalizadas.

Es importante recordar, como se ha mencionado anteriormente, que esta etapa consta de dos fases: Fase I y Fase II.

5.5.1 Dispersión del Proceso

El enfoque para monitorizar la dispersión del proceso se basa en la varianzas generalizadas de la muestra, $|S|$, que es el determinante su matriz de varianzas-covarianzas. Por ejemplo, un método consiste en utilizar la esperanza y la varianzas de $|S|$, es decir, $E(|S|)$ y $V(|S|)$, y la propiedad de que la mayor parte de la distribución de probabilidad de $|S|$ se encuentra en el intervalo $E(|S|) \pm 3\sqrt{V(|S|)}$, para determinar el límite de control superior e inferior para monitorear la dispersión del proceso (Roberto José Herrera Acosta 2008).

La extensión multivariante de los gráficos de control para la dispersión del proceso no es tan directa como la del promedio del proceso. Un modelo general y técnicas que abarquen una amplia gama de problemas encontrados en la práctica no están disponibles. En la mayoría de los casos, los problemas particulares deben abordarse de manera específica.

5.5.2 Desplazamiento de la Media

Para el detectar un posible desplazamiento de la media con respecto del valor objetivo se utilizan los gráficos de control tipo Shewhart. Dichos gráficos usan información de la muestra actual y son relativamente insensibles a pequeños o moderados desplazamientos de la media. Gráficos de control multivariados como MEWMA se utilizan para detectar desplazamientos de 0.5 a 2σ más rápido puesto que se usan medias ponderadas.

5.5.2.1 T^2 de Hotelling

Una vez comprobada la normalidad de nuestras observaciones $N_d(\mu_0, \Sigma_0)$, para la fase I, cuando una serie de vectores observados en características multivariantes, Tabla 1, se han obtenido y se quiere utilizar gráficos de control multivariados de tipo Shewhart para monitorear el proceso industrial, se estiman el vector de media de la muestra \bar{x} y el vector de varianza-covarianza de la muestra S . La observación j -ésima, no es independiente de \bar{x} ni de S .

Estado	Estadístico	UCL
fase I $n = 1,$	$T_j^2 = (X_j - \bar{x})^T S^{-1}(X_j - \bar{x})$	$U_{CL} = \frac{(m-1)^2}{m} B_{1-\alpha, d/2, \left(\frac{2(m-1)^2}{3m-4} - d-1\right)/2}$
fase II $n = 1,$	$T_j^2 = (X_f - \bar{x})^T S^{-1}(X_f - \bar{x})$	$U_{CL} = \frac{d(m+1)(m-1)}{m(m-d)} F_{1-\alpha, d, m-d}$
fase I $n > 1,$	$T_j^2 = n(\bar{x}_j - \bar{\bar{x}})^T \bar{S}^{-1}(\bar{x}_j - \bar{\bar{x}})$	$U_{CL} = \frac{d(m-1)(n-1)}{mn-m-d+1} F_{1-\alpha, d, mn-m-d+1}$
fase II $n > 1,$	$T_f^2 = n(\bar{x}_f - \bar{\bar{x}})^T \bar{S}^{-1}(\bar{x}_f - \bar{\bar{x}})$	$U_{CL} = \frac{d(m-1)(n-1)}{mn-m-d+1} F_{1-\alpha, d, mn-m-d+1}$

Tabla 2 Resumen expresión estadístico y límite de control por fase

Esta tabla recoge el cálculo de la secuencia de puntos para construir el gráfico de control, así como el cálculo del límite de control, que en el caso de que alguna observación sea mayor que el límite, que nos indica que no tenemos control sobre nuestro proceso productivo y que habría que realizar algunos cambios.

En nuestro caso vamos a representar la fase I para $n = 1$ puesto que por las especificaciones de nuestro problema no tenemos subgrupos racionales. Utilizamos la herramienta Minitab para ello.

5.5.2.2 MEWMA

Las gráficas de control EWMA se desarrollaron para proporcionar mayor sensibilidad a pequeños cambios en el caso univariante, y se pueden extender a problemas de control de calidad multivariante. Al igual que en el caso univariante, la versión multivariante (MEWMA) de esta gráfica siempre se utiliza como procedimiento en la fase II.

Nuevamente, en esta etapa de la metodología se calcula una secuencia de estadísticos para la construcción de un gráfico pero que, como se ha mencionado, aporta mayor sensibilidad a la hora de detectar desplazamientos ligeros de la media.

Un análisis de ARL de la gráfica de control MEWMA se puede utilizar para seleccionar los parámetros λ y h basados en la simulación, considerando el tamaño del subgrupo, el tamaño del desplazamiento, el *ARL* en control dado ARL_0 , y el rendimiento del *ARL* correspondiente a un tamaño del desplazamiento dado. Se elige h para lograr un ARL_0 , especificando en control. λ se selecciona para minimizar el *ARL* para un tamaño del desplazamiento de la media $\delta = [(\mu - \mu_0)^T \Sigma^{-1} (\mu - \mu_0)]$ como se muestra a continuación en la tabla.

Shift size δ	0,5	1	1,5 to 2	3
λ	0,03	0,1	0,15 to 0,25	0,4

Tabla 3 valores de Lambda para desplazamiento sigma

Esta tabla nos indica los valores óptimos para λ para detectar los desplazamientos δ .

5.6 Caracterización de la Población

Una vez estudiadas las observaciones y de haber refinado nuestro proceso mecánico (puesto que hemos eliminado atípicos {5} y en los gráficos de control hemos identificado y eliminado las causas asignables de variación {6}) es el momento, desde una perspectiva industrial, de inferir y de determinar cómo se distribuyen nuestros datos de forma teórica [7]. Esto nos va a servir para saber cómo es la población de procedencia de nuestros datos para comprobar si existe algún aspecto que se ha pasado por alto, es decir, estimar los parámetros para utilizarlos en futuros métodos, puesto que el proceso industrial continuará cuando se acabe el estudio (en este caso de taladros en chapa metálica), para identificar nuevos aspectos a tener en cuenta (como posibles grupos) nos basamos en métodos estadísticos como: MLE y clustering (kmedias).

En este apartado no se van a obtener resultados porque su utilidad queda fuera del alcance de este trabajo.

5.6.1 Estimación de los Parámetros

En esta etapa se estimarán los parámetros necesarios, sin embargo, las estimaciones se pueden obtener del apartado de métodos descriptivos numéricos.

5.6.2 Clasificación de las Observaciones en familias

Clasificar las observaciones en clústeres nos ayuda a encontrar nuevos factores que en este estudio no se han tenido en cuenta para el futuro.

Se distinguen principalmente dos tipos de algoritmos: los basados en la “reubicación iterativa” y los basados en la “agrupación jerárquica”, sin embargo, los últimos son más apropiados para observaciones categóricas por lo que en este estudio nos centraremos en el primer tipo.

La reubicación iterativa parte de un número determinado de clústeres en los que mediante criterios de distancias encuentra la semejanza óptima entre las observaciones. El método que se va a utilizar es kmedias.

El método de las k-medias consiste en la agrupación de las observaciones a la partición cuyo centroide es el más cercano a una observación determinada según un determinado criterio de distancias (Manuel Castejón Limas, Joaquín Ordieres Meré, Francisco Javier de Cos Juez, Francisco Javier Martínez de Pisón 2014).

5.7 Índices de Capacidad del Proceso

El propósito de un índice de capacidad de proceso es reflejar cuán bien o cuán mal un proceso productivo genera productos conformes con las especificaciones. El uso del índice de capacidad de proceso para cantidades multivariadas refleja mejor este comportamiento del proceso industrial que para múltiples cantidades univariantes (ISO 22514-6).

Una evaluación multivariante será requerida cuando se cumpla al menos una de las siguientes circunstancias:

- Los límites de tolerancia no se pueden expresar independientemente para todas las características, por ejemplo, cuando el límite de una característica es función de otra (condición de máximo material).
- Las características que pueden ser combinadas están correlacionadas entre sí.

En este trabajo se describen dos tipos de índices: Tipo I y Tipo II cuya diferencia radica en si el índice está basado en probabilidades (Tipo I) o definido geoméricamente entre áreas o volúmenes (o hiper-volúmenes) de la región de tolerancia del proceso (Tipo II).

- I. Tipo I: Está basado en la probabilidad P de conforme o no conforme, el índice se calcula usando las relaciones entre el índice y la probabilidad dicha entre el caso normal univariante.
- II. Tipo II: El índice se calcula con el ratio entre el área o volumen (o hiper-volumen) de la región de tolerancia y la región de variación del proceso.

Para poder comparar los intervalos de variación del proceso con la región de tolerancia es necesario aplicar ciertas transformaciones que se describen a continuación.

- a) Transformar la región de tolerancia a la forma del intervalo de variación del proceso.
- b) Transformar el intervalo de variación del proceso en la forma de la región de tolerancia.
- c) Transformar ambas regiones a una nueva dimensión orientada en una función.

Si al concluir con esta parte se determina que el proceso industrial no es capaz (es decir, que el $C_p < 1$), entonces hay una serie de medidas que se pueden adoptar, pero en cualquiera de los casos hay que redefinir las actuaciones preparatorias (1). Las medidas que se pueden adoptar son: 1) Centrar el proceso. 2) Mejorar el proceso (con una máquina de proceso mecánico mejor, por ejemplo). 3) Evaluar si las especificaciones son muy restrictivas y se admite su modificación. 4) Continuar con el mismo proceso, pero realizando el 100% de las inspecciones. 5) Finalizar la producción.

Por otra parte, si nuestro C_p es $1 \leq C_p \leq 1.33$, la medidas apropiadas a efectuar son: 1) Controlar el centrado del proceso. 2) Volver a localizar la media del proceso para así conseguir un $C_{pk} = C_p$. 3) Continuar monitorizando los gráficos de control.

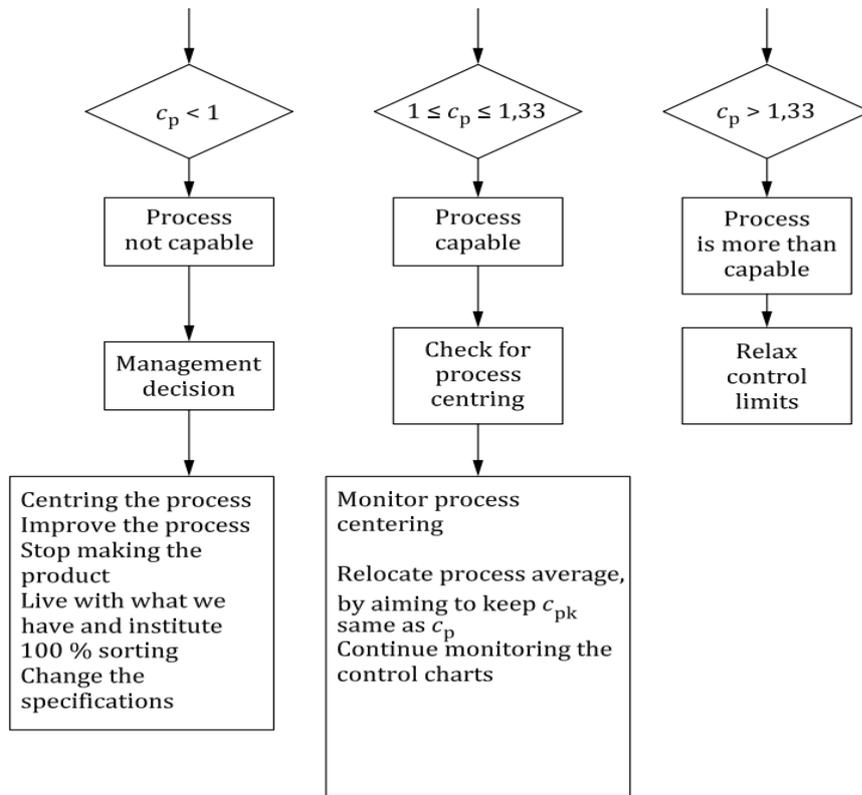


Figura 18 Consideraciones C_p

En la Figura 18 se puede observar las diferentes medidas a tomar en función del resultado para el C_p .

5.7.1 Índice de Capacidad Tipo Ia:

Considerando la distribución normal d-dimensional. Si la región de tolerancia de los casos no es de forma elíptica (círculo o elipse si $d = 2$, esfera o elipsoide si $d = 3$, hiper-esfera o hiper-elipsoide si $d > 3$) entonces se transformará a una región que sí que tenga esa forma. Esto se hace determinando la máxima elipse (o elipsoide o hiper-elipsoide) centrada en el valor nominal y que esté completamente contenida en la región de tolerancia original.

Para calcular el índice C_p , la distribución normal debe estar centrada en el centro de la región de tolerancia elíptica. Para esta distribución normal, hay que determinar el contorno elipsoidal que esté completamente contenido en la región de tolerancia elíptica y calcular la probabilidad del volumen limitado por ese contorno elipsoidal bajo la distribución normal d-dimensional. Denotar como P la probabilidad, entonces el índice C_p se calcula cómo:

$$C_p = \frac{1}{3} \Phi^{-1} \left(\frac{P + 1}{2} \right)$$

El valor de C_p necesario variará según los requisitos del cliente o incluso según las normas de referencia vigentes en el momento (generalmente se necesita un $C_p > 1.33$) pero, el \widehat{C}_p tendrá un valor de 1 cuando $\widehat{P} = 99.73\%$, cuanto más grande sea la probabilidad, mayor será el valor del \widehat{C}_p . Esto está explicado en el Anexo A.

El cálculo del C_{pk} involucra la media y la varianza de la distribución (porque en este caso no se centra el intervalo de variación del proceso) y se calcula de 2 posibles formas:

- El contorno máximo que está completamente contenido en la región de tolerancia transformada, si μ esta dentro de la región de tolerancia.
- El contorno máximo que no está contenido en la región de tolerancia transformada, si μ está fuera de la región de tolerancia.

$$C_{pk} = \frac{1}{3} \Phi^{-1} \left(\frac{P+1}{2} \right) \quad \text{si } \mu \text{ esta dentro de la región de tolerancia}$$

$$C_{pk} = \frac{1}{3} \Phi^{-1} \left(\frac{1-P}{2} \right) \quad \text{si } \mu \text{ está fuera de la región de tolerancia}$$

El valor de C_{pk} necesario variará según los requerimientos del cliente o incluso según las normas de referencia vigentes en el momento (generalmente se necesita un $C_{pk} > 1.33$) pero, el \widehat{C}_{pk} tendrá un valor de 1 cuando $\widehat{P} = 99.73\%$, cuanto más grande sea la probabilidad, mayor será el valor del \widehat{C}_{pk} . Esto está explicado en el Anexo A.

El Tipo Ia descrito anteriormente es el que se utiliza para calcular la capacidad en nuestro problema. Este índice encuentra aplicaciones en el dimensionamiento geométrico y la tolerancia de desviaciones de posición. En este caso, la nueva región de tolerancia describe una zona de tolerancia circular. Los símbolos que se utilizan a menudo en esos casos son C_p y C_{pk} .

5.7.2 Índices de Capacidad Tipos Ic y Iic.

Los índices de capacidad de proceso Tipo Ic y Tipo Iic se basan en una transformación que simplifica la caracterización de múltiples características en una sola. En este enfoque, se expresa el aspecto multivariado mediante la función de transformación $q(x)$, donde x describe la cantidad multivariada. Esta transformación debe reflejar la importancia funcional de las cantidades individuales en x y su interacción. Por ejemplo, se puede considerar como un modelo para la región de tolerancia y se interpreta como una función de ponderación o una función de cuantificación que evalúa la funcionalidad técnica.

En primer lugar, debemos definir una función de evaluación técnica, denotada como $q(x)$, que se aplique a la región de tolerancia d -dimensional. Esta función alcanza su valor máximo, q_{max} en el objetivo dentro de dicha región. En el límite de la región de tolerancia, $q(x)$ toma el valor q_{bound} . En algunos casos, podemos derivar los valores de q_{max} y q_{bound} a partir del contexto técnico de todas las cantidades individuales en x . En otros casos, utilizamos valores estándar, como $q_{max} = 1$ y $q_{bound} = 0,5$. La función $q(x)$ puede expresarse mediante una ecuación cerrada o como una composición de funciones.

En el segundo paso, las piezas producidas deben ser muestreadas y medidas. Luego, los valores medidos se transforman mediante $q(x)$ en valores de calificación relacionados con la función.

En el tercer paso, a partir de estos valores, se debe identificar una función de distribución univariante apropiada, denotada como $F(q)$ que mejor ajuste nuestros datos. Como alternativa, se puede realizar una segunda transformación a una distribución normal univariante. Si la función de calificación $q(x)$ aumenta de manera monótona desde el límite hasta el objetivo, y el vector aleatorio X sigue una distribución normal multivariante, la densidad de distribución de $F(q)$ será unimodal.

En el cuarto paso, basado en la distribución identificada para los valores de calificación, el objetivo transformado y los límites de especificación, se calculan los índices de capacidad. Si se eligen $q_{max} = 1$ y $q_{bound} = 0,5$, 0,5 nos da el límite de especificación inferior y 1 el límite superior natural. Como q sólo da un intervalo de tolerancia unilateral y una distribución límite unilateral, se puede usar únicamente el C_{pk} para realizar el análisis.

Si el índice es de Tipo I, C_{pk} se calcula análogamente a 5.7.1 siendo P la resultante de la distribución ajustada $F(q)$ y $P = 1 - F(q_{bound})$.

Si el índice es de Tipo II, definiendo X_{mid} como la mediana de los $F(q)$, y definiendo L como q_{bound} y $\Delta L = X_{mid} - X_{0,135\%}$, la ecuación es:

$$C_{pk} = \frac{q_{50\%} - q_{bound}}{q_{50\%} - q_{0,135\%}} \quad \text{donde } q_{x\%} \text{ es el } x\% \text{-percentil de } F(q)$$

5.7.3 Índices de Capacidad Tipos IIa y IIb

Los índices de capacidad de Tipo II siguen el principio de relacionar y extienden la región de tolerancia a un el extremo de la región de variación del proceso. Estos extremos se expresan en áreas o volúmenes. El área o volumen de la región de tolerancia se denota por V_{tol} y V_{proc} denota el área o volumen de la región de proceso. Por lo que el índice se puede definir cómo:

$$C_p = \left[\frac{V_{tol}}{V_{proc}} \right]^a$$

El exponente a se introduce para dar la posibilidad de reducir el área o volumen a una dimensión. Por lo que es normalmente $a = 1/d$. Si no, $a = 1$.

5.7.3.1 Índice de Capacidad Tipo IIa

Nuevamente, se transforma la región de tolerancia a la elipse (o elipsoide o hiper-elipsoide) de mayor tamaño que está completamente contenido en los límites de tolerancia originales. En el caso de la región de tolerancia donde los parámetros $x_{t1}, x_{t2}, \dots, x_{td}$ denotan las medias distancias de los límites de tolerancia al valor nominal/objetivo, entonces, el volumen del proceso V_{tol} es:

$$V_{tol} = \frac{\pi^{d/2}}{\Gamma(1 + \frac{d}{2})} \prod_{i=1}^d x_{ti}$$

Para estimar el C_p , el volumen del elipsoide 99,73% se estima como

$$V_{proc} = \frac{\pi \chi_{99.73\%}^{2 \cdot d/2}}{\Gamma(1 + \frac{d}{2})} \sqrt{|S|}$$

Este índice es apoyado por el valor de $\frac{1}{D}$ que da información acerca de la localización del proceso en relación con el valor nominal/objetivo μ_0 . D se estima cómo:

$$\hat{D} = \sqrt{1 + \frac{n}{n-1} (x - \mu_0)^T S^{-1} (x - \mu_0)}$$

Y combinados se estima el C_{pm}

$$\hat{C}_{pm} = \frac{V_{tol}}{V_{proc}} \cdot \frac{1}{\hat{D}}$$

5.7.3.2 Índice de Capacidad Tipo IIb

La forma del intervalo de variación del proceso, que es elíptica, se transforma a la región de tolerancia. En el caso de que la región de tolerancia tenga forma rectangular (o

cúbica o hiper-cúbica), la región de variación del proceso se transforma a un rectángulo (o cubo o hiper-cubo) de menor tamaño que encapsule totalmente a la región de tolerancia y basado en las proyecciones de los intervalos de la elipse de cada dimensión (Se calculan los índices como univariantes con la región transformada).

Como el valor del C_{pk} es habitualmente inferior a valor del C_p , es usual utilizar para la interpretación del conocimiento de tu proceso únicamente el C_{pk} , sin embargo no hay que rechazar el C_p , puesto que un resultado válido para el C_p cuando el C_{pk} es malo nos puede indicar que tu proceso, aunque tenga precisión alta, la exactitud no es adecuada, por lo que únicamente habría que calibrar la máquina (Juan José Saavedra, Josselyn Gautier, Amalio monzón, Silvia Pascual, M^a Ángeles Martín, Félix González, Carolina López Egea Gómez, Felix Isabel, Nora Galavis 2023).

6 Resultados

En este capítulo se van a mostrar los resultados obtenidos de aplicar las fases mencionadas a lo largo de este estudio al conjunto de datos de la Tabla 1.

En la Figura 4 se observa la nube de puntos de la Tabla 1 y en rojo el valor teórico de la posición de los taladros.

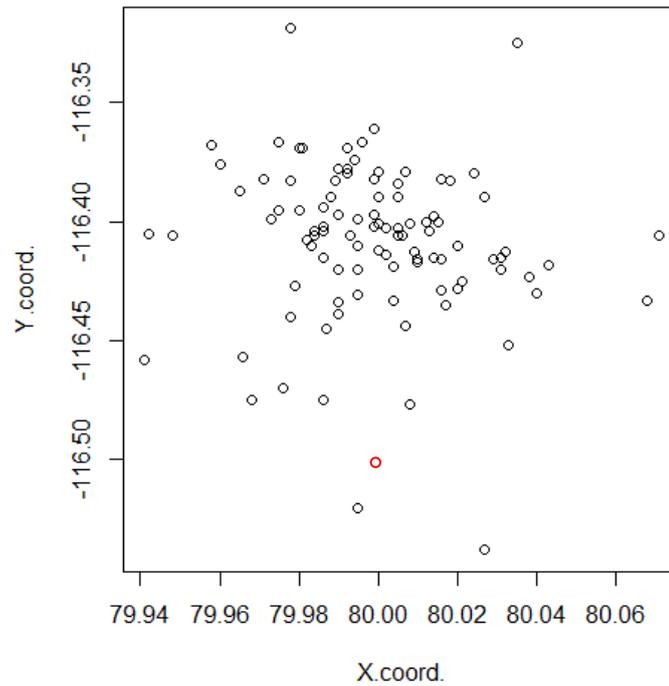


Figura 4 Diagrama de puntos de las observaciones y valor nominal (punto rojo)

De la nube de puntos de la Figura 5 se puede ver como la media de las observaciones para el eje Y es claramente superior al valor nominal del plano, además, los puntos se encuentran distribuidos de forma parecida en ambos ejes, es decir las varianzas en el eje X e Y son comparables (aunque una inspección más precisa nos indicará que esto no es verdad)

Para realizar el análisis de componentes principales se ha utilizado la matriz de covarianzas puesto que ambas variables son coordenadas (posición en el eje X e Y) y por lo tanto tienen la misma escala y se han obtenido los siguientes resultados:

```
eigen() decomposition
$values
[1] 0.0010868556 0.0005260283

$vectors
      [,1]      [,2]
[1,] -0.1349617 -0.9908508
[2,]  0.9908508 -0.1349617
```

El primer autovector nos indica la dirección de mayor variabilidad de nuestros datos y la segunda componente viene dada por la componente ortogonal. Además, el tamaño de cada autovalor con respecto a la suma de los dos autovalores nos indica la importancia relativa de cada autovector, es decir el porcentaje de variabilidad que nos aporta esa dirección dada por el autovector.

La información preservada (ip_1) por la primera componente es:

$$ip_1 = 100 \cdot \frac{0.0010868556}{0.0010868556 + 0.0005260283} = 67,3858\%$$

La información preservada (ip_2) por la segunda componente es:

$$ip_2 = \frac{0.0005260283}{0.0010868556 + 0.0005260283} = 32,6142\%$$

El análisis descriptivo se ha llevado a cabo obteniendo como medidas numéricas las que se presentan a continuación y realizando diagramas de cajas e histogramas

	X. coord.	Y. coord.
Min.	:79.94	Min. :-116.5
1st Qu.	:79.99	1st Qu. :-116.4
Median	:80.00	Median :-116.4
Mean	:80.00	Mean :-116.4
3rd Qu.	:80.01	3rd Qu. :-116.4
Max.	:80.07	Max. :-116.3

En este caso se está haciendo una interpretación univariante de nuestro caso de estudio. Donde se puede observar el rango, los cuartiles, la media y la mediana.

```
> cov(datos) #matriz de varianzas-covarianzas de los datos
      X.coord.      Y.coord.
X.coord. 5.362435e-04 -7.499768e-05
Y.coord. -7.499768e-05 1.076640e-03
> cor(datos) #matriz de correlación de los datos
      X.coord.      Y.coord.
X.coord. 1.00000000 -0.09870325
Y.coord. -0.09870325 1.00000000
```

Se realizan ahora los gráficos mencionados anteriormente sobre los datos de la Tabla 1.

Para la Posición del eje X:

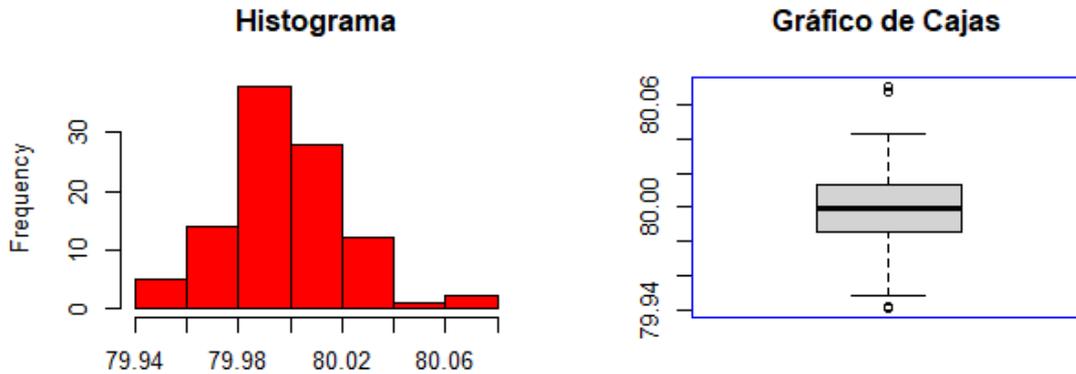


Figura 5 Resumen Eje X de las observaciones

Para la Posición en el eje Y:

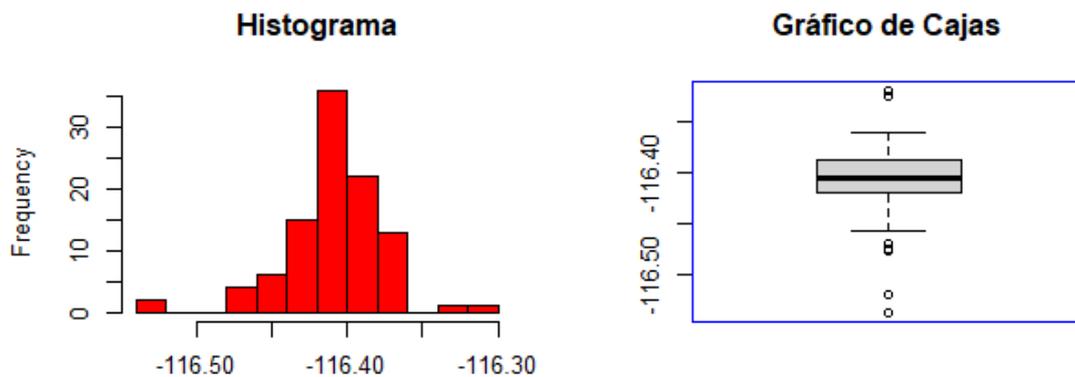


Figura 6 Resumen Eje Y de las observaciones

Para los gráficos de cajas de la Figura 6 y Figura 7, aunque aparentemente la longitud de la caja en la posición del Eje X es mayor, si nos fijamos en la escala nos damos cuenta de que el rango para la posición del eje Y es mayor, además, se pueden observar más potenciales observaciones atípicas en el diagrama para la posición Y (puntos más alejados de los “bigotes”) por lo que se puede concluir que la varianza de la máquina de mecanizado para el eje Y va a ser mayor que para el eje X.

Para los gráficos de cajas de la Figura 6 y Figura 7, el histograma nos indica como el rango de las observaciones para el eje Y es mayor, adicionalmente, se aprecia como hay más observaciones concentradas en torno al centro del rango (puesto que la altura del histograma es mayor) sin embargo, como he mencionado anteriormente, existen potenciales espúreos en las observaciones próximas a los extremos del rango.

Del análisis de los datos de la Tabla 1 se observa que la máquina no trabaja con la misma precisión en el eje X que en el eje Y, que sería algo para tener en cuenta a la hora de realizar nuevos taladros. Esto es importante para el control de proceso y la necesidad de que los servicios técnicos intervengan en el mantenimiento del equipo. Es una conclusión del análisis exploratorio de las distribuciones marginales y se trata de algo que hay que comprobar en todos los estudios similares.

El análisis de normalidad se observa en los gráficos cuantil-cuantil de la figura 7 que se ha explicado anteriormente, se utilizan para evaluar la normalidad de las observaciones de una forma gráfica.

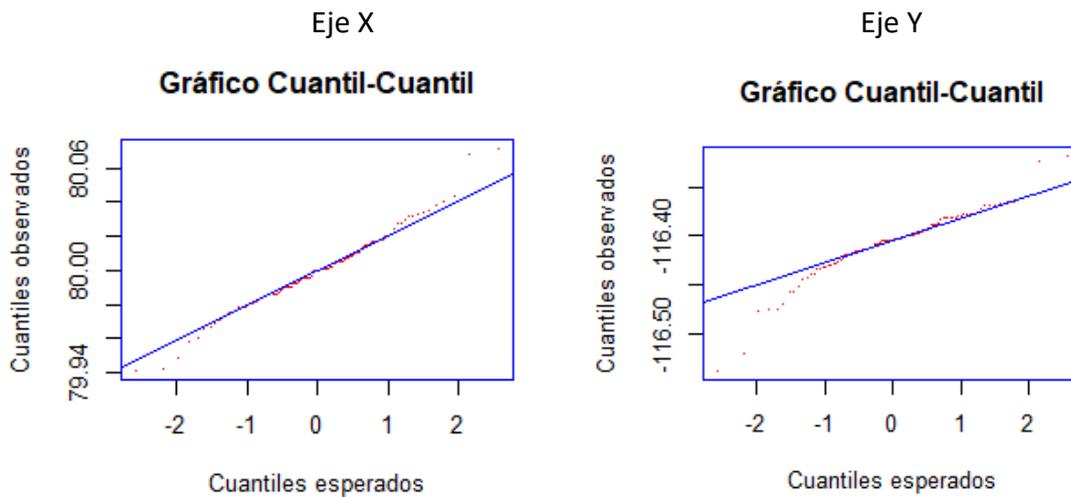


Figura 7 Gráfico cuantil-cuantil

De la Figura 7 no se puede concluir con que los datos de la Tabla 1 se aproximen a una distribución normal.

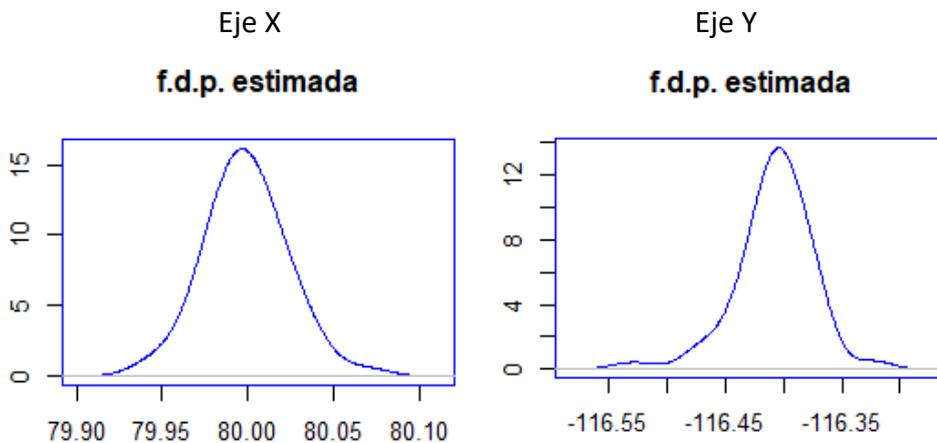


Figura 8 Función de probabilidad estimada para cada variable

Mientras que en los gráficos cuantil-cuantil Figura 8 no se puede apreciar nada concluyente, es verdad que en los gráficos de función de probabilidad estimada si que se aproxima a una curva normal, aunque en las colas es posible que existan atípicos.

Asimismo, los tests de Mardia, llevados a cabo con las observaciones de la Tabla 1, nos permite evaluar la normalidad para los valores de asimetría y kurtosis de nuestros datos.

Para el primer coeficiente:

```
asimetria <- function(data.in){
  N <- nrow(data.in); n <- ncol(data.in)
  mij3 <- 0
  invS <- solve(cov(data.in))
  for (i in 1:N)
    for (j in 1:N)

      mij3 <- mij3 + (data.in[i,] %% invS %% data.in[j,])^3
  xalpha0.01 <- qchisq(0.99, n*(n+1)*(n+6)/6)
  return (list(N_b1n_6=mij3/(N*6), xalpha0.01=xalpha0.01 ))
}
```

```
datos.b1n <- asimetria(datos)
unlist(datos.b1n )
#se ve que el valor del estadístico está fuera
#de h0 para una alfa del 99%,
#porque el valor del estadístico observado es
#mayor que el percentil 99 de la distribución
#chi-cuadrado
```

```
      N_b1n_6  xalpha0.01
1.852961e+23 2.009024e+01
```

Esto no se trata de un contraste de hipótesis al uso, lo que se calcula es el valor del estadístico observado y se compara con el percentil determinado de la distribución que sigue el estadístico. Como el valor observado es mayor a el valor del percentil se puede afirmar que a nivel 1% se rechaza H_0 , además, como el valor observado es mucho más grande que el percentil se puede decir que el p-valor asociado al contraste va a aproximarse mucho a 0.

Para el segundo coeficiente:

```

kurtosis <- function(data.in){
  n <- ncol(data.in)
  N <- nrow(data.in)

  mahadatos <- mahalanobis(data.in, apply(data.in, 2, mean), cov(data.in))
  xalpha0.01 <- qnorm(0.99, n*(n+2), sqrt(8*n*(n+2)/N))
  return(list(b2n = sum(mahadatos^2)/N))
}

n <- 2
N <- 100
xalpha0.01 <- qnorm(0.99, n*(n+2), sqrt(8*n*(n+2)/N))

datos.kurtosis <- kurtosis(datos)
c(unlist(datos.kurtosis),xalpha0.01)

```

b2n
11.894073 9.861078

Nuevamente se comprueba que el valor observado del estadístico queda fuera del IC del 99% asique se rechaza la normalidad. Además, aunque este estadístico calculado es más robusto para muestras grandes (que no es nuestro caso habitualmente) por lo que tiene menos importancia, en este caso el p-valor asociado no será tan pequeño al del contraste anterior, pero sigue siendo inferior a 0.01.

De una forma computacionalmente más sencilla se analiza la normalidad de las componentes principales puesto que está demostrado que si cada una de las componentes siguen una distribución normal, la distribución conjunta será normal.

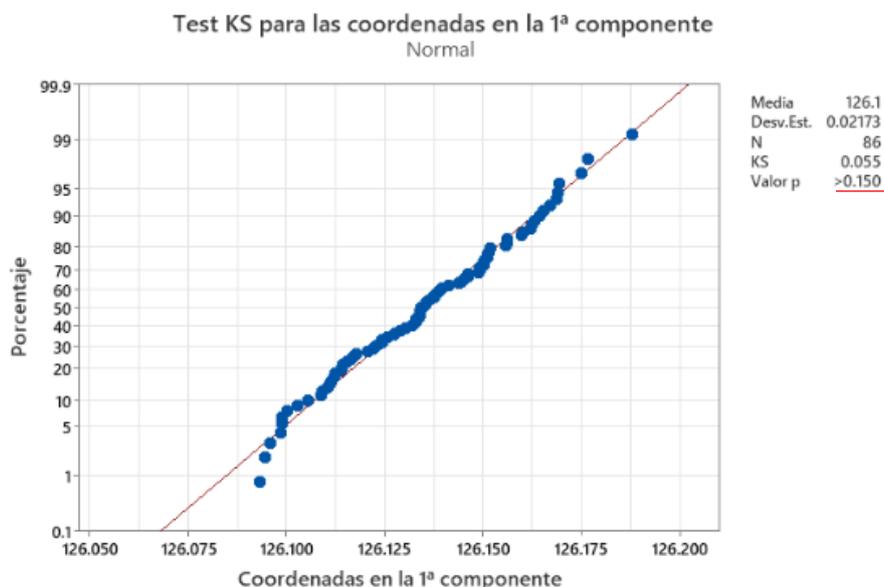


Figura 9 Cálculo de la normalidad KS para las coordenadas en la primera componente

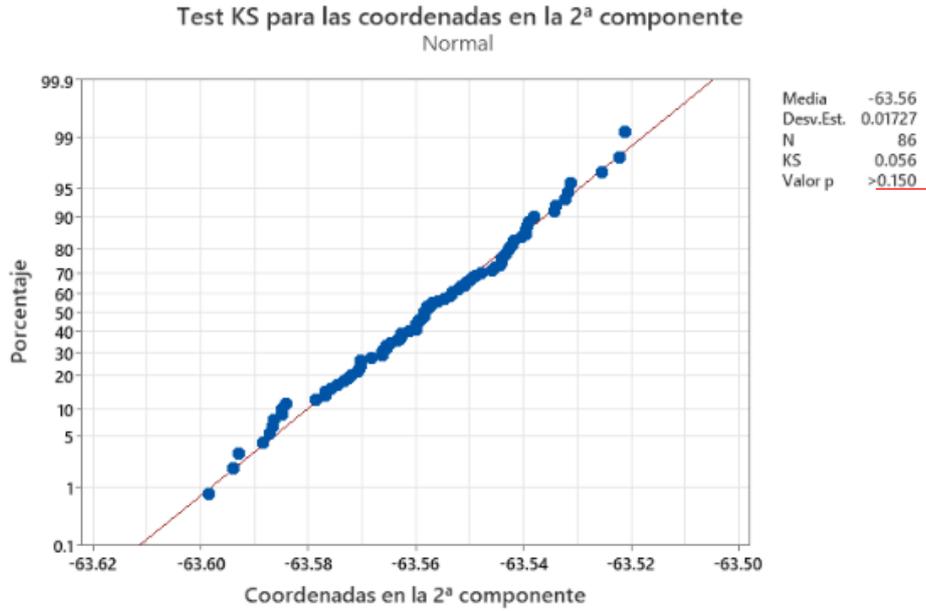


Figura 10 Cálculo de la normalidad KS para las coordenadas en la segunda componente

En la Figura 9 y Figura 10 se puede comprobar que las coordenadas de las observaciones de la Tabla 1 en las componentes principales siguen una distribución normal univariante, por lo que la distribución conjunta de los valores originales (Tabla 1) sigue una distribución normal bivariante.

Los valores atípicos se detectan mediante las distancias de Mahalanobis puesto que algunas observaciones pueden quedar ocultas en la nube de puntos siendo una observación atípica si se tiene en cuenta que su distribución de partida es multivariada.

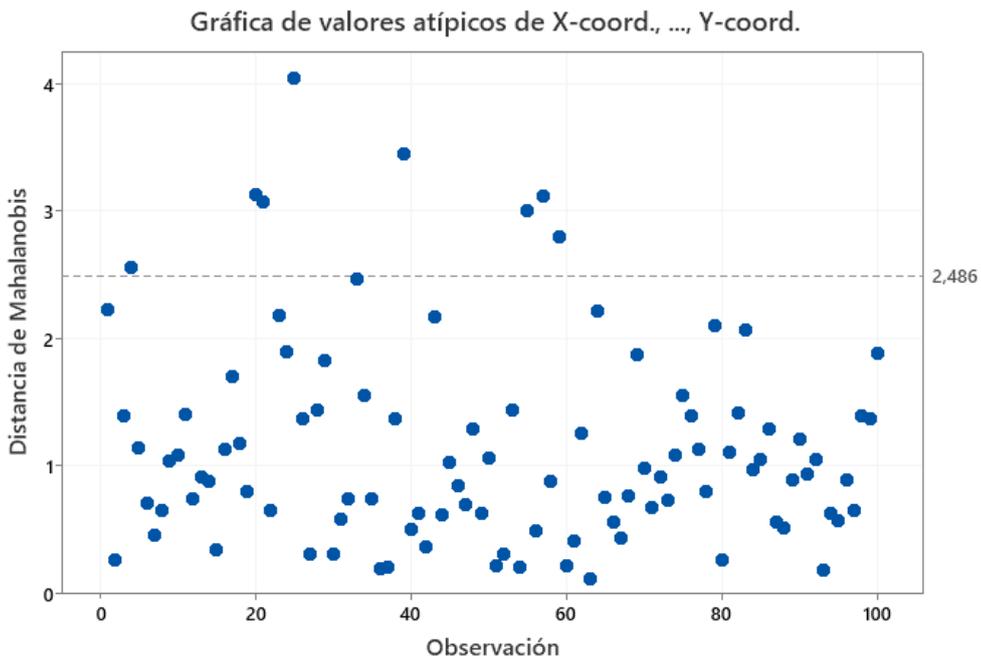


Figura 11 Gráfico para valores atípicos

La distribución de las distancias de Mahalanobis nos dice que las observaciones 4, 20, 21, 25, 39, 55, 57 y 59 se tratan de observaciones atípicas, puesto que su valor observado del estadístico es superior al percentil-99 (2,486 acompañado de línea discontinua de la Figura 11), de la distribución que siguen. Es muy conveniente identificar las causas que generan estos datos para posteriormente eliminarlas.

Realizamos nuevamente esta prueba para la detección de observaciones atípicas con los nuevos datos sin los *outliers* obtenidos anteriormente.

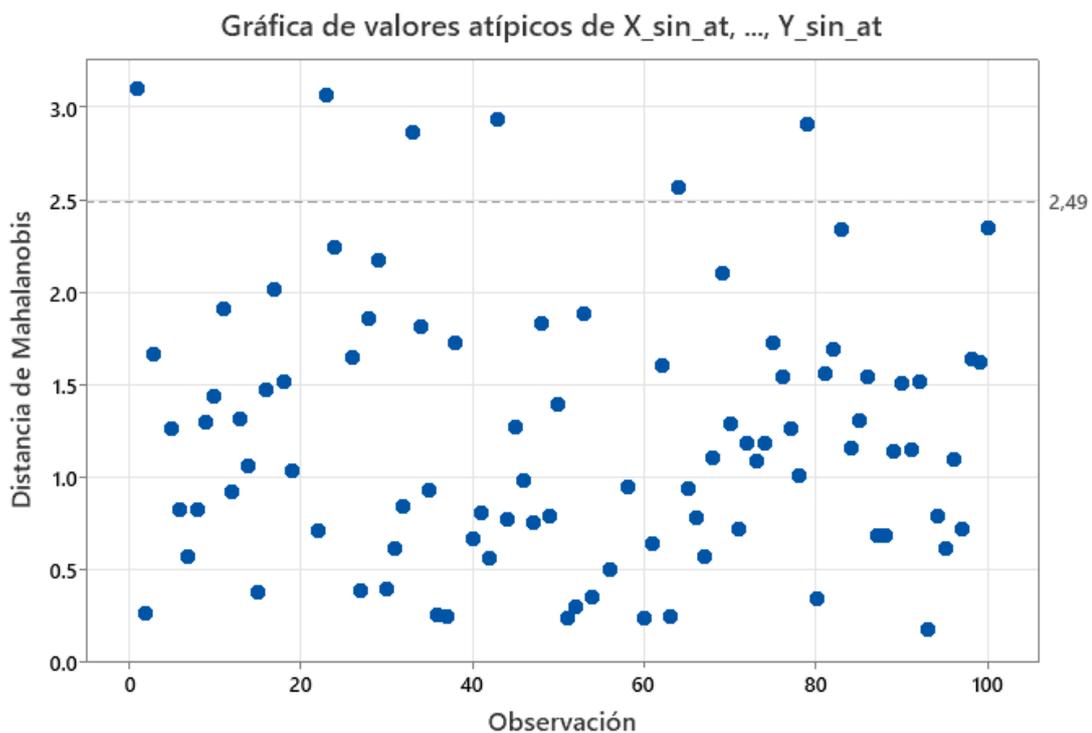


Figura 12 Gráfico para valores atípicos

En esta nueva iteración del método indica la existencia de nuevos atípicos, que son las observaciones 1, 23, 33, 43, 64 y 79 puesto que su valor observado del estadístico es superior al percentil-99 (2,49 acompañado de línea discontinua de la Figura 12). Nuevamente, tras estudiar las causas que los provocaron, se procede a su eliminación.

En la siguiente iteración del método existen observaciones que se consideran *borderline*, no habiéndose identificado causas asignables, se aceptan como valores de la distribución y se da por finalizada las iteraciones del método.

Si se comprueba la normalidad de los datos, una vez finalizado el tratamiento de los datos, se observa como los p-valores asociados a los contrastes nos quedan, para las componentes, de 0.783 para la coordenada X y 0.536 para la coordenada Y por lo que se asume que ha habido algún suceso fuera de lo común a la hora de fabricación de esas observaciones atípicas (vibraciones/otro operario/...).

Los gráficos de control son necesarios para saber si el proceso es estable, en caso de que no lo sean mandarían una señal de fuera de control. Estos gráficos se han obtenido con la herramienta Minitab, calculando para cada gráfico de control su correspondiente secuencia de estadísticos, cada uno basado en el suyo, estas secuencias han sido explicadas anteriormente.

En la Figura 13 Se presentan los resultados proporcionados por Minitab para el gráfico de varianza generalizada. En primer lugar se realiza con las muestra de la Tabla 1 al completo y posteriormente (Figura 14) tras el tratamiento de atípicos. Esta doble presentación para enfatizar el valor de la eliminación de los atípicos.

- Gráfico para la muestra completa:

Gráfico de Varianza Generalizada con Atípicos

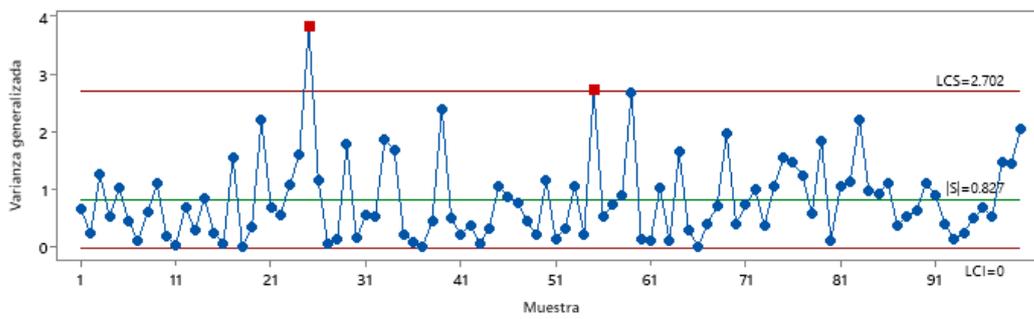


Figura 13 Gráfico de control de varianza generalizada con atípicos

En la Figura 13 se aprecia como para la observación 25 nos queda un valor muy alto, por encima del límite de control superior (LCS) que nos indica que puede haber pasado algo en esa observación. Para la observación número 55 también se observa un valor elevado pero muy próximo al límite de control. Sin embargo, necesitaríamos más información para tomar decisiones acerca del proceso industrial. Esta es la gráfica para los valores originales de la Tabla 1, es decir, sin quitar las observaciones atípicas. Con los valores sin observaciones atípicas nos queda:

- Gráfico para la muestra sin atípicos:

Gráfico de Varianza Generalizada sin Atípicos

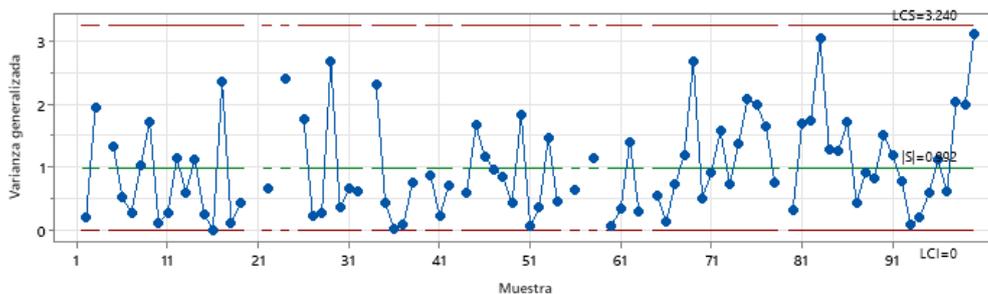


Figura 14 Gráfico de control de varianza generalizada sin atípicos

Donde ahora, Figura 14, no se percibe ninguna observación por encima del límite de control superior.

En el gráfico de control T^2 se representa la secuencia mencionada anteriormente. El límite de control representa valores atípicos de este estadístico y nos indica que en esa observación se ha desplazado la media (o el valor teórico) a la hora de realizar el taladro lo que conlleva que el taladro no esté tan bien localizado como debería.

- Gráfico para la muestra completa:

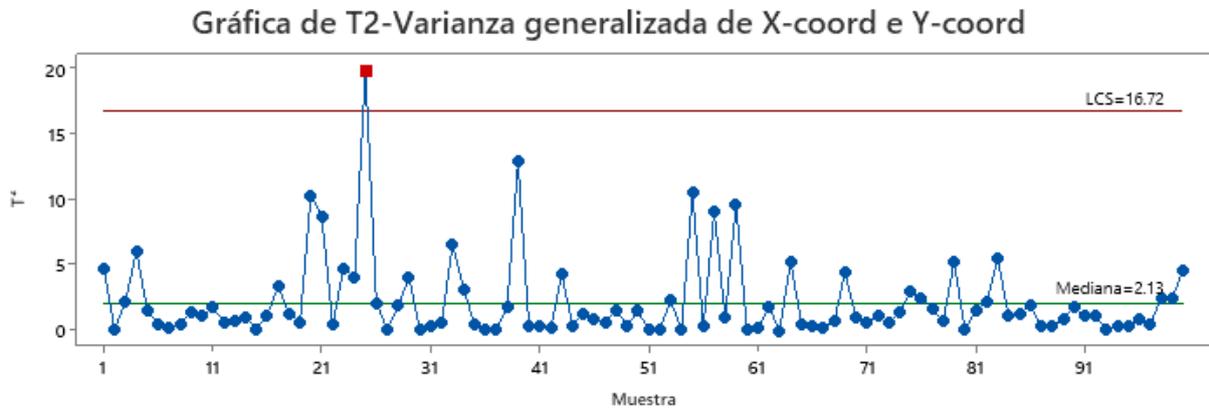


Figura 15 Gráfico de control T^2 de Hotelling con atípicos

Nuevamente empiezo con la gráfica para los valores de la Tabla 1, sin quitar las observaciones atípicas.

- Gráfico para la muestra sin atípicos:

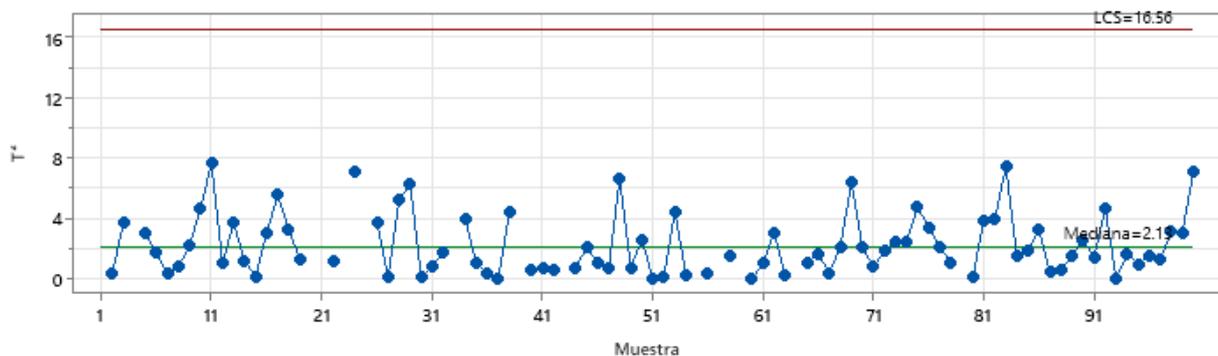


Figura 16 Gráfico de control T^2 de Hotelling sin atípicos

Los valores observados en la Figura 16 están por debajo del límite de control por lo que no hay razones para dudar de la validez de nuestro proceso industrial.

El gráfico de control MEWMA da un fuera de señal cuando detecta desplazamiento de la media, pero de una forma más sensible que el gráfico de control de T^2 , estos gráficos se utilizan de manera conjunta puesto que este gráfico puede dar un fuera de señal y que no sea necesario efectuar ningún cambio en el proceso.

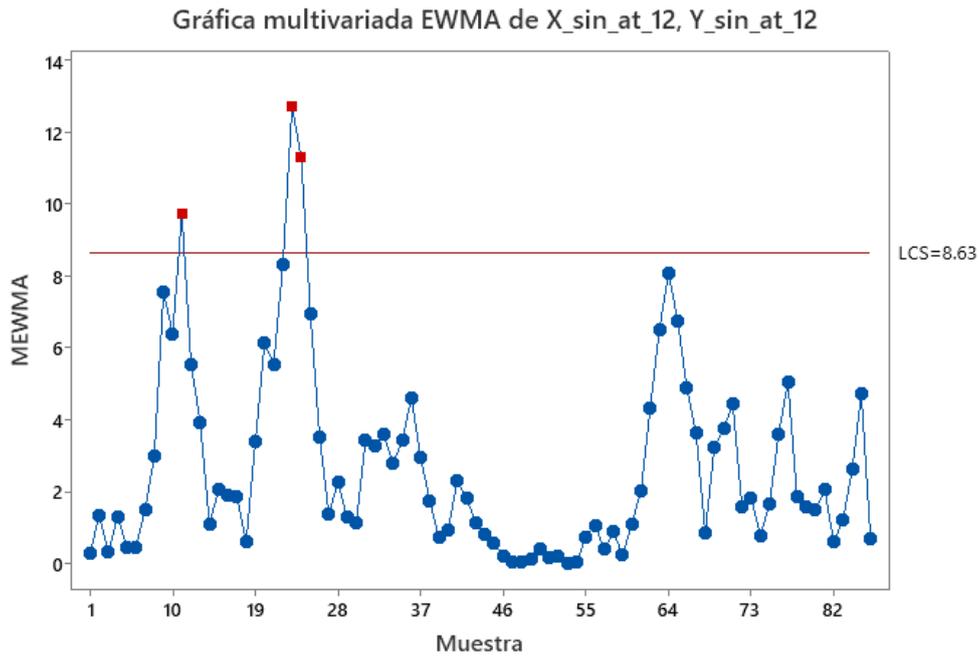


Figura 17 Gráfico de control MEWMA

Los valores para los que el gráfico, Figura 17, nos da la señal de fuera de control son los que están relacionados con los valores atípicos eliminados anteriormente, por lo que observaciones que no han sido fabricados inmediatamente seguidos se interpretan como si lo hubiesen sido, además, los valores posteriores quedan dentro de los límites por lo que proseguimos con el estudio. Esto ha tenido que ser necesario porque Minitab para este método no permite valores *na* (not available) y se ha creado un nuevo conjunto de datos con todas las observaciones seguidas. Aun así, este gráfico nos detecta pequeños desplazamientos de la media por lo que, en el caso de que nos diera alguna observación fuera de control, aunque habría que tomar alguna medida la observación sería válida.

En esta fase final de determinar la capacidad del proceso se va a poder observar la relación entre el intervalo de variación del proceso y el intervalo de tolerancia de las especificaciones de plano. Lo que se busca es que el intervalo de tolerancia sea mucho más pequeño que el de variación, lo que conlleva un C_p y un C_{pk} mayores que 1.

```
X <- data.frame(datosISO8.1.1[,3])
Y <- data.frame(datosISO8.1.1[,4])

datos <- as.matrix(data.frame(c(X, Y)))

mu <- c(80, -116.5)
medias <- (c(mean(X), mean(Y)))

options(digits = 22)

#para calcular el Cp trabajamos con los valores de plano, por lo que el radio es 0.25
r <- 0.25

chi2_val <- r^2/max(eigen(cov(datos))$values)

probabilidad <- pchisq(chi2_val, df=2)
print(probabilidad)

Cp <- (qnorm((probabilidad+1)/2))/3; Cp

#para calcular el Cpk trabajamos con la media de nuestro proceso
#el nuevo radio se calcula obteniendo el punto de corte de la circunferencia
# de la nueva región de tolerancia con la recta que pasa por el punto de la media del
#proceso, en nuestro caso (79.9992, -116.4082) y que tiene por pendiente la dirección
# de la primera componente

r2 <- 0.1587

chi2_val2 <- r2^2/max(eigen(cov(datos))$values)

probabilidad2 <- pchisq(chi2_val2, df=2)
print(probabilidad2)

Cpk <- (qnorm((probabilidad2+1)/2))/3; Cpk
```

```

> #para calcular el Cp trabajamos con los valores de plano, por lo que el radio es 0.25
>
> r <- 0.25
>
> chi2_val <- r^2/max(eigen(cov(datos))$values)
>
> probabilidad <- pchisq(chi2_val, df=2)
> print(probabilidad)
[1] 0.99999999999967426
>
> Cp <- (qnorm((probabilidad+1)/2))/3; Cp
[1] 2.4277683614545849
>
> #para calcular el Cpk trabajamos con la media de nuestro proceso
> #el nuevo radio se calcula obteniendo el punto de corte de la circunferencia
> # de la nueva región de tolerancia con la recta que pasa por el punto de la media del
> #proceso, en nuestro caso (79.9992, -116.4082) y que tiene por pendiente la dirección
> # de la primera componente
>
> r2 <- 0.1587
>
> chi2_val2 <- r2^2/max(eigen(cov(datos))$values)
>
> probabilidad2 <- pchisq(chi2_val2, df=2)
> print(probabilidad2)
[1] 0.99999070927022482
>
> Cpk <- (qnorm((probabilidad2+1)/2))/3; Cpk
[1] 1.477684888258801

```

Como se puede observar, tanto el C_p como el C_{pk} dan valores superiores a 1.33 por lo que se puede decir que el proceso es capaz. Como se ha observado en la Figura 18 las actuaciones a realizar ahora sería relajar los límites de control, además, se está en disposición de empezar a hablar de reducción de inspecciones.

7 Conclusiones

La necesidad de seguir esta metodología de forma meticulosa se debe a la industria en la que nos encontramos puesto que los resultados de tomar conclusiones erróneas pueden ser catastróficos (como un accidente de aviación), además, esta industria también conlleva dificultades añadidas puesto que no es habitual tener muestras con muchas observaciones (lo que es normal porque si se compara, por ejemplo, el número de coches que hay con el número de aviones que hay se llega a la conclusión de que las cadencias de producción son bajas).

En la etapa de descripción de las observaciones observamos como la máquina no trabajaba igual en el eje X y en el eje Y por lo que hay que tomar medidas al respecto, como es calibrar la máquina o en el caso que no se pueda, comprar una nueva.

Al haberse demostrado la normalidad de los datos, se puede determinar que las observaciones atípicas se deben a fuentes de variación no asignables, es decir, como no se puede decir la procedencia de la variación "atípica", el tratamiento más apropiado es la eliminación de los atípicos

También se ha demostrado con los gráficos de control que el proceso es estable, puesto que los gráficos de control no nos dan observaciones fuera de los límites de control, como ya se ha mencionado anteriormente, únicamente dan señales de fuera de control a observaciones que no debería ir seguidas pero que se trabaja como si lo fueran puesto que el software utilizado lo requiere. Que el proceso esté bajo control nos indica que la forma en la que se mitigan las fuentes variación de conocidas es apropiada, porque no existe señales de

fuera de control (que nos indicaría que en la fabricación de esa pieza ha ocurrido un suceso anómalo que altera el valor de la observación).

En cuanto a la capacidad del proceso, se ha comprobado que el proceso es capaz. Esto quiere decir que la relación entre la varianza de la distribución de probabilidad del resultado de las piezas y entre los límites de especificación (límites de tolerancias del plano de la pieza) es un número grande (el intervalo de variación es mucho más pequeño que el intervalo de tolerancia) por lo que, como se ha mencionado antes, se está en disposición de poder relajar los límites de control o, incluso, poder hablar sobre una posible reducción de inspecciones.

8 Bibliografía

(2017). ISO 7870-7, Control charts — Part 7: Multivariate control charts.

(2017). ISO 22514-2, Statistical methods in process management — Capability and performance — Part 2: Process capability and performance of time-dependent process models.

(2013). ISO 22514-6, Statistical methods in process management — Capability and performance — Part 6: Process capability statistics for characteristics following a multivariate normal distribution.

MONTGOMERY, Douglas C. (2005). Introduction to statistical quality control. Wiley, 6th ed.

Manuel Castejón Limas, Joaquín Ordieres Meré, Francisco Javier de Cos Juez, Francisco Javier Martínez de Pisón. (2014). Control de Calidad Metodología para el análisis previo a la modelización de datos en procesos industriales. Fundamentos teóricos y aplicaciones prácticas con R. Universidad de La Rioja.

(2024). AS 13003, Measurement Systems Analysis Requirements for the Aero Engine Supply Chain.

Ralph B. D'Agostino, Michael A. Stephens. (1986). Goodness-of-Fit-Techniques.

Jobson. (1994). Multivariate Data Analysis. Volume II.

Ana Justel Daniel Peña and Rubén Zamar. (1997). A MULTIVARIATE KOLMOGOROV-SMIRNOV TEST OF GOODNESS OF FIT.

(2013). ISO 3534-2, Statistics — Vocabulary and symbols — Part 2: Applied statistics.

Roberto José Herrera Acosta. (2008). CONTROL CHART T-CUADRADO - INDEX CAPABILITY MULTIVARIATE.

Juan José Saavedra, Josselyn Gautier, Amalio monzón, Silvia Pascual, M^a Ángeles Martín, Félix González, Carolina López Egea Gómez, Felix Isabel, Nora Galavis. (2023). 2023 Guía Core Tools de APQP.

Daniel J. Denis. (2010). Applied-Univariate-Bivariate-And-Multivariate-Statistics-Using-Python.

(2013). UNE-ISO 3534-2, Estadística Vocabulario y símbolos. Parte 2: Estadística Aplicada.

(2005). ISO 9000 Sistemas de gestión calidad.

(2021). ISO 2692 Especificación geométrica de productos. Requisito máximo material (MMC).

(2018). ASME Y-14.5 Dimensioning and tolerancing Standard.

(2017). ISO 1101 Especificación geométrica de productos. Tolerancia geométrica y de forma.

(1963). AMERICAN STANDARDS INSTITUTE. 5Proposed American Standard flowchart symbols for information processing Comm ACM 6, 10, 601-604.

(1970). ANSI Standard flowchart symbols and their use information processing (X3.5) American National Standards Institute, New York, (20 pp.).

Stephans, M.A. (1974). "EDF Statistics for Goodness of Fit and Some Comparisons", Journal of American Statistical Association 69, 730-737.

(2017). ISO 7870-2, Control charts — Part 2: Shewhart control charts.

9 Anexos

9.1 Anexo A: Definiciones

[2], [3] y [1]

Característica: Rasgo distintivo de una pieza fabricada o un proceso de producción.

Características multivariantes: Cantidad multivariante donde una serie de características diferentes consisten en d cantidades que están solas o combinadas con la calidad de un producto.

Nota 2: La observación de características multivariantes puede ser expresada como el vector $x = (x_1, x_2, x_3, \dots, x_d)^T$. Por lo que una característica multivariante puede ser considerada como una característica vector de un producto. El valor de la cantidad multivariante se representa como un punto en el espacio d-dimensional.

Nota 3: Todas las cantidades individuales pueden ser medidas en el mismo producto u objeto.

Nota 4: Si la cantidad multivariante se describe como una media de estadísticos, el vector se considera como un vector aleatorio que sigue una distribución d-dimensional.

Región de confianza: Intervalo d-dimensional para características multivariantes d-dimensionales y definida para un nivel de confianza determinado.

Nota 1: Esta región está limitada por líneas, superficies o hiper-superficies en el espacio d-dimensional y no necesariamente cerrado.

Nota 2: La forma y el tamaño de la región de confianza está definida por uno o más parámetros.

Región de Tolerancia: Región en el espacio de la característica que contiene todos los valores permitidos para la característica multivariante.

Nota 1: Esta región está limitada por líneas, superficies o hiper-superficies en el espacio d-dimensional y no necesariamente cerradas.

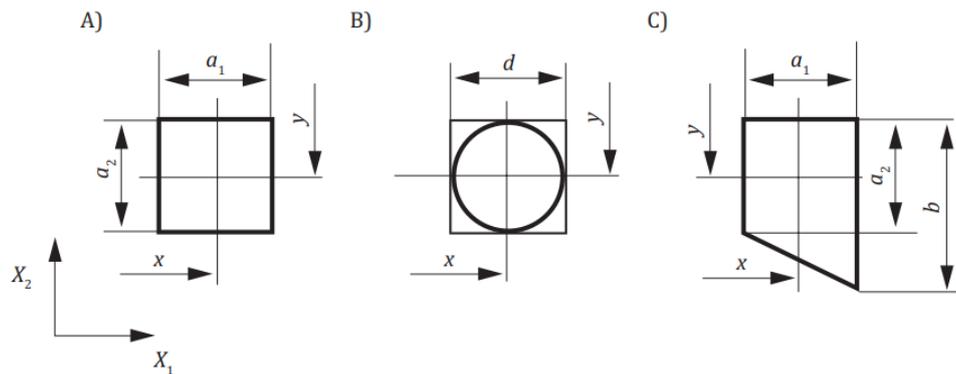
Nota 2: La forma y el tamaño de la región de tolerancia está definida por uno o más parámetros.

Nota 3: Las formas típicas de la región de tolerancia son rectángulos, elipses (o círculos) en el caso bidimensional, cuboides o hiper-cuboides, elipsoides o hiper-elipsoides o formas prismáticas compuestas en el caso tridimensional.

Nota 4: La región de tolerancia se especificada en el plano técnico que describe las condiciones geométricas de la pieza. Los productos que posean valores fuera de la región de

tolerancia se asumen que no cumplen los requerimientos funcionales necesarios. Estos productos se consideran piezas no conformes.

Nota 5: Para evaluar un producto con respecto a los límites de región de tolerancia, el orden de las características individuales en la característica multivariante y el número d de dimensiones debe ser igual al que describe la región de tolerancia.



A Región de tolerancia rectangular con parámetros a_1 , a_2 , x e y .

B Región de tolerancia circular con parámetros d , x e y .

C Región rectangular extendida triangularmente con parámetros a_1 , a_2 , b , x e y .

Proceso: Conjunto de actividades que se suceden y que transforma elementos de entrada en elementos de salida.

Nota 1: Los elementos de entrada de un proceso son generalmente de salida de otros procesos.

Nota 2: Los procesos de una organización generalmente se planifican y se llevan a cabo bajo condiciones controladas para añadir un valor adicional.

[16] y [15]

Capacidad de Proceso: La distribución de los valores de una característica medida en un proceso, del que previamente se ha demostrado que está bajo control estadístico. Describe la habilidad de ese proceso para producir valores para esa característica que cumpla los requerimientos para esa característica.

Nota 1: Algunos ejemplos representativos son: El espesor de una pared, el diámetro de un taladro, la temperatura de un baño de limpieza, la posición de una ranura respecto de un sistema de referencia (datum), ...

Nota 2: La abreviatura genérica empleada en este trabajo para el índice de capacidad del proceso es PCI.

Rendimiento del proceso: Distribución de los valores de las características medidas de un proceso.

Nota 1: El proceso puede no estar demostrado que está bajo control estadístico.

[15]

Subgrupo racional: Subgrupo en el que la variabilidad se supone debida solamente a causas aleatorias. Es una denominación que permite identificar la forma en la que se selecciona la muestra evaluada y aplicarla en los gráficos de control.

Nota 1: Un subgrupo es un conjunto de datos tomados de un proceso de modo que asegure la mayor similitud entre los datos de cada subgrupo y la mayor diferencia entre los datos de subgrupos diferentes. Cuanto mayor es un subgrupo, más sensible es el gráfico de control a cambio en el nivel de proceso. De manera ideal, cada medida en un subgrupo es independiente de otras.

Nota 2: El método más común para obtener un subgrupo racional consiste en formarlo en un momento dado. Los datos de periodos de tiempo distintos estarían por tanto en subgrupos distintos. Un ejemplo sería tomar medidas de cinco piezas consecutivas de una máquina en particular cada hora. Los estadísticos muestrales de los subgrupos pueden representarse en un gráfico de control ordenado por tiempos. Esto facilita la detección de variaciones relacionadas con el tiempo.

[15]

Gráfico de control: Gráfico en el que se presentan los valores de una medida estadística sobre una serie de muestras en un orden particular para hacer el seguimiento del proceso respecto a esa medida y para controlar y reducir la variabilidad.

Nota 1: El orden particular generalmente se basa en el tiempo o en el número de orden de la muestra.

Nota 2: El gráfico de control es más eficaz cuando la medida es una variable del proceso que está relacionada con el producto final o una característica de servicio.

9.2 Anexo B: Términos Abreviados

9.2.1 Acrónimos

MMC: Condición de máximo material (maximum material condition).

PCI: Índice de capacidad del proceso (process capability index).

SPC: Control de proceso estadístico (statistical process control).

MSPC: Control de proceso estadístico multivariante (multivariate statistical process control).

ACP (o PCA): Análisis de componentes principales (principal components analysis).

PLS: suma de cuadrados parciales (partial squared sum).

LCS (o UCL): Límite de control superior (upper control limit).

LCI (o LCL): Límite de control inferior (lower control limit).

ARL: Tiempo de proceso medio (Average run length).

EWMA: Media móvil ponderada exponencial (exponentially weighted moving Average).

MEWMA: Media móvil ponderada exponencial multivariante (multivariate exponentially weighted moving Average).

ANOVA: Análisis de la varianza.

MANOVA: Análisis multivariado de la varianza.

MLE: estimador de máxima verosimilitud (Maximum likelihood estimation).

RFT: Pieza buena a la primera (Right First Time).

9.2.2 Símbolos

$B_{1-\alpha, \nu_1, \nu_2}$: el cuantil $1 - \alpha$ para la distribución beta con ν_1 y ν_2 grados de libertad.

d : número de dimensiones para características multivariantes.

D_j^2 : estadístico graficado en el gráfico de control para la fase II χ^2 .

$E(|S|)$: media teórica de $|S|$.

$F_{1-\alpha, \nu_1, \nu_2}$: el cuantil $1 - \alpha$ para la distribución F con ν_1 y ν_2 grados de libertad.

h : límite de control superior del gráfico de control MEWMA.

L_{CL} : límite de control inferior.

m : número de subgrupos.

n tamaño de cada subgrupo.

N : número de observaciones.

$N_d(\mu, \Sigma)$: distribución normal d-dimensional con μ y Σ .

S_{ab} : covarianza entre las características a y b con $n = 1$.

S_{abj} : covarianza entre las características a y b en el subgrupo j con $n > 1$.

S_i^2 : varianza de la característica i con $n = 1$.

S_{ij}^2 : varianza de la característica i en el subgrupo j con $n > 1$.

\bar{s}_i^2 : media de S_{ij}^2 sobre todos los subgrupos para la característica i con $n > 1$.

\bar{s}_{ab} : media de los s_{ab} para todos los subgrupos para la covarianza entre las características a y b para $n > 1$.

S : matriz de varianzas-covarianzas muestral para $n = 1$.

\bar{S} : matriz de varianzas-covarianzas muestral para $n > 1$.

$|S|$: determinante de la matriz de varianzas-covarianzas muestral S .

T_j^2 : estadístico graficado en la fase I.

T_f^2 : estadístico graficado en la fase II.

tr : operador traza.

U_{CL} : límite de control superior.

$V(|S|)$: varianza de $|S|$.

X_{ij} : la observación j -ésima para la característica i -ésima con $n = 1$.

X_{ijk} : la observación k-ésima para el subgrupo j-ésimo y la característica i-ésima con $n > 1$.

\bar{x}_{ij} : media de la característica i-ésima en el j-ésimo subgrupo con $n > 1$.

x : vector de observaciones.

x_j : vector de características para la observación j-ésima con $n = 1$.

x_f : vector de características para una observación futura con $n = 1$.

\bar{x} : vector de medias para $n = 1$.

\bar{x}_j : vector medias para el subgrupo j-ésimo con $n > 1$.

\bar{x}_f : vector medias para un subgrupo futuro con $n > 1$.

$\bar{\bar{x}}$: vector de medias muestrales para $n > 1$.

\vec{x} : vector d-dimensional.

$\{\bar{\bar{x}}_i\}$: elemento i-ésimo del vector $\bar{\bar{x}}$.

Y_j^2 : estadístico graficado en el gráfico de control MEWMA.

Z_j : estadístico MEWMA.

$\chi_{1-\alpha, \nu}^2$: cuantil $1 - \alpha$ para la distribución χ^2 con ν grados de libertad.

δ : vector de desplazamiento de la media.

λ : vector de parámetro móvil de MEWMA.

λ : parámetro móvil de EWMA, con $0 < \lambda \leq 1$.

μ : vector medias teóricas para las características multivariantes.

μ_0 : vector medias teóricas pre-especificado para las características multivariantes.

ρ_{y_1, y_2} : coeficiente de correlación entre y_1 e y_2 .

Σ : matriz de varianzas-covarianzas teóricas para características multivariantes.

Σ_0 : matriz de varianzas-covarianzas teóricas pre-especificadas para características multivariantes.

Σ_{Z_j} : matriz de varianzas-covarianzas para el estadístico MEWMA Z_j .

Λ : matriz de varianzas-covarianzas diagonalizada en la nueva base.

$(\cdot)^{-1}$: operador inversa.

$(\cdot)^T$: operador transpuesta.

C_p : índice de capacidad de proceso.

C_{pk} : índice de capacidad mínima del proceso.

C_{pkL} : índice de capacidad inferior del proceso.

C_{pkU} : índice de capacidad superior del proceso.

P_p : índice de rendimiento del proceso.

P_{pk} : índice de rendimiento mínimo del proceso.

P_{pkL} : índice de rendimiento inferior del proceso.

P_{pkU} : índice de rendimiento superior del proceso.

9.3 Anexo C

La densidad de la distribución normal d-dimensional con media μ y matriz de covarianzas Σ es:

$$f(x) = \frac{1}{2\pi \left(\frac{d}{2}\right) |\Sigma| \left(\frac{1}{2}\right)} \exp\left(-\left(\frac{1}{2}\right) (x - \mu)^T \Sigma^{-1} (x - \mu)\right), \quad x \in R^d,$$

Si la matriz Σ es definida positiva y tiene inversa. Aquí X y μ son vectores d-dimensionales y Σ es una matriz $d \times d$. Los vectores son vectores columna y T denota la transpuesta de un vector o una matriz. Esta distribución normal d-dimensional se denota como $N_d(\mu, \Sigma)$.

Los contornos de densidad constantes son:

$$\{x | (x - \mu)^T \Sigma^{-1} (x - \mu) = c^2\}$$

Y son intervalos para $d = 1$, elipses para $d = 2$ y elipsoides para $d \geq 3$.

La probabilidad de que el valor de un proceso esté dentro del área limitada por el contorno de las elipses puede ser calculada con la distribución χ^2 con d grados de libertad. Si X sigue una distribución $N_d(\mu, \Sigma)$, entonces

$$P((x - \mu)^T \Sigma^{-1} (x - \mu) \leq c^2) = F_{\chi^2(d)}(c^2)$$

Donde $F_{\chi^2(d)}$ denota la función de distribución de la χ^2 con d grados de libertad.

Por lo que el contorno de la elipse sigue

$$\{x | (x - \mu)^T \Sigma^{-1} (x - \mu) = \left(\sqrt{F_{\chi^2(d)}^{-1}(p)}\right)^2\}$$

Es el límite de una región con probabilidad p . Aquí, $F_{\chi^2(d)}^{-1}(p)$ es el p-cuantil de la distribución χ^2 con d grados de libertad.

