

Sistema de familias de distribuciones de Johnson, una alternativa para el manejo de datos no normales en cartas de control

INGRID JULIANA LAGOS^{*}
JOSE ALBERTO VARGAS^{**}

Resumen

La carta de control para datos individuales o Carta X , es una herramienta estadística útil en procesos donde no es posible tomar muestras de tamaño mayor a 1, sin embargo, esta presenta problemas cuando el supuesto de normalidad sobre el cuál están construidas no se satisface, ya que asumir normalidad cuando no se está seguro de ella puede llevar a tomar decisiones erróneas como declarar un proceso fuera de control cuando en realidad está bajo control incrementando el riesgo tipo I o por el contrario incrementando el riesgo tipo II cuando se declara un proceso bajo control estadístico y no lo está. Una alternativa es hacer uso de transformaciones que permitan llevar los datos a una distribución normal a través del Sistema de Familias de Distribuciones de Johnson, que consiste en determinar la distribución (dentro de la familia de distribuciones de Johnson) a la que pertenecen los datos, y una vez determinada, realizar transformaciones sobre ellos, con el fin de poder aplicar la teoría clásica de las Cartas de Control. Como resultado del estudio de esta metodología, se llega a mostrar que el Sistema de Familias de Distribuciones de Johnson es una herramienta muy útil cuando se desea encontrar una transformación que logre normalizar un conjunto de datos.

Palabras Clave: Carta X , No-Normalidad, Sistema de Familias de Distribuciones de Johnson.

^{*}Estadística de la Universidad Nacional de Colombia. E-mail: julianalagos@yahoo.com.mx

^{**}Profesor asociado, Departamento de Estadística, Universidad Nacional de Colombia. E-mail: avargas@matematicas.unal.edu.co

1. Introducción

Una herramienta estadística frecuentemente usada dentro del Control Estadístico de Procesos es la Carta de Control, instrumento útil para el estudio y el monitoreo de la estabilidad de un proceso a través de una característica de calidad, que permite detectar patrones anormales en cualquier momento.

En la práctica se desea controlar el valor promedio de la característica de calidad así como su variabilidad. Estos controles se hacen a través de cartas de control para el monitoreo del nivel medio del proceso, en donde se utilizan medidas de tendencia central como la media (Cartas \bar{X}) y la mediana (Cartas \tilde{X}), y cartas de control para el monitoreo de la dispersión del proceso, en donde se utilizan medidas como la desviación estándar (Cartas S) y los rangos (Cartas R). Estos tipos de cartas son las más usadas en el control estadístico de procesos en línea.

La carta \bar{X} es la más popular cuando se desea controlar el nivel medio del proceso. Esta consiste en tomar muestras de tamaño n ($n \geq 2$) a intervalos igualmente espaciados y ubicar las medias de estas muestras dentro de un gráfico conformado por tres líneas paralelas a saber: límite de control superior (LCS), línea central (LC) y límite de control inferior (LCI). Si las medias se encuentran dentro de la región conformada por los límites superior e inferior, la variación del proceso se debe a causas aleatorias propias del proceso. Por el contrario si alguno de esos puntos se encuentra fuera de la región, la variación del proceso se debe a causas atribuibles, lo cual exige una inmediata intervención a fin de determinar la causa y tomar medidas correctivas, llevando el proceso de nuevo a control.

Sin embargo en muchas situaciones sólo se pueden tomar muestras de tamaño $n = 1$ en cada tiempo, característica propia de procesos continuos o en línea donde la recolección de datos se hace a través de maquinaria sofisticada que permite acumular grandes cantidades de información en cualquier momento, procesos donde la tasa de producción es lenta o también aquellos en donde las mediciones repetidas difieren únicamente por errores de laboratorio o de análisis, donde es muy frecuente encontrar distribuciones no normales. Montgomery (2001).

Para el manejo de datos provenientes de distribuciones no-normales en SPC, se han desarrollado diferentes métodos alternativos. Uno conocido como el Sistema de Familias de Distribuciones de Johnson, (Chou, Polansky & Mason 1998), que consiste en determinar la distribución (dentro de la familia de distribuciones de Johnson) a la que pertenecen los datos, y una vez determinada, realizar transformaciones sobre ellos, con el fin de poder aplicar la teoría clásica

de las Cartas de Control. Este sistema es conocido en el ámbito de control de calidad por brindar mayor flexibilidad en la transformación de un conjunto de datos a distribución normal. El presente trabajo se centra en el control del nivel medio de este tipo de procesos, a través de la Carta \bar{X} ó Carta para Datos Individuales. Esta carta presenta problemas cuando el supuesto de normalidad no se satisface, ya que asumir normalidad cuando no se está seguro de ella puede llevar a tomar decisiones erróneas como declarar un proceso fuera de control cuando en realidad está bajo control (incremento del Riesgo tipo I), generar índices de capacidad altos cuando la capacidad real es baja, aumentar la tasa de rechazos, etc, elevando los costos de producción, algo bastante desfavorable en la industria cuando se trata de una falsa alarma.

En la sección 2 de este documento se define la carta de control para datos individuales o carta \bar{X} , en la sección 3 se hace una introducción al Sistema de Familias de Distribuciones de Johnson, en la sección 4 se presenta una aplicación al sector industrial de las geomembranas de PVC y finalmente en la sección 5 se dan algunas conclusiones.

2. Carta de control para datos individuales

Existen casos en donde los datos no pueden conformar grupos de tamaño $n \geq 2$, tales como procesos en los que la tasa de producción es lenta, o aquellos que cuentan con maquinaria sofisticada que permite la medición segundo a segundo de la característica de calidad bajo estudio, es decir una observación en cada tiempo t . Para este tipo de procesos en donde adicionalmente se desea monitorear el nivel medio del proceso, es adecuada la construcción de una Carta \bar{X} para observaciones individuales que en adelante se denotará simplemente como Carta \bar{X} . Sea X_1, X_2, \dots, X_n el conjunto de n observaciones de la característica de calidad bajo estudio asumidas normales con media μ y varianza σ^2 . Se define $R_i = |X_i - X_{i+1}|$ como el valor absoluto del i -ésimo rango móvil para $i = 1, 2, 3, \dots, n - 1$. La carta \bar{X} queda definida de la siguiente manera:

$$LCS = \bar{X} + \frac{3\bar{R}}{d_2} \quad (1)$$

$$LC = \bar{X} \quad (2)$$

$$LCI = \bar{X} - \frac{3\bar{R}}{d_2}, \quad (3)$$

donde $\bar{X} = \sum_{i=1}^n X_i/n$, $\bar{R} = \sum_{i=1}^{n-1} |R_i|/(n - 1)$ y d_2 es una constante que depende del tamaño de las amplitudes móviles es decir de $n = 2$. Esta constante

se encuentra tabulada en diferentes textos, ver por ejemplo Montgomery (2001), Mittag (1993).

Para evaluar la eficiencia de una carta de control se define la Longitud de Corrida (Run Length), RL , como el número de puntos graficados en una carta de control hasta que aparezca la primera señal fuera de control. Dado que este valor cambia de ensayo a ensayo, se define la Longitud Promedio de Corrida (Average Run Length), ARL , como el valor esperado de la variable aleatoria RL . Teóricamente una carta de control bajo características ideales de normalidad tiene un ARL de 370, Vargas (2000). Sin embargo, se ha comprobado que ante desvios del supuesto de normalidad (distribuciones asimétricas) este valor disminuye dramáticamente mostrando que la aplicación de esta carta no es adecuada, pues se llega a incrementar el riesgo tipo I o II definidos previamente en la sección 1. (Lagos 2002). Para el manejo de datos no normales en control de calidad se han propuesto diversas metodologías, una de ellas es el Sistema de Familias de Distribuciones de Johnson que se presenta a continuación.

3. Sistema de familias de distribuciones de Johnson

Cuando el supuesto de normalidad sobre la característica de calidad bajo estudio no se cumple, se presentan problemas para la aplicación de algunas técnicas en Control Estadístico de Procesos (Statistical Process Control), SPC. Sin embargo, cuando esto sucede es posible transformar los datos no-normales a datos normales, a través de técnicas como el Sistema de Familias de Distribuciones de Johnson.

Johnson en el año de 1949 define tres familias de distribuciones para una variable aleatoria X continua a saber:

- S_B : Se refiere a X acotada.
- S_L : Se refiere a X acotada por debajo o lognormal.
- S_U : Se refiere a X no-acotada.

Para ajustar un conjunto de datos no-normales, es necesario establecer criterios que permitan determinar la pertenencia del mismo a una de las tres familias. Cada una de ellas tiene asociada una transformación de X a una variable normal estándar Z , así como condiciones especiales para los parámetros estimados y el rango de la variable X , que deben tenerse en cuenta cuando se va a escoger la

familia con la que se quiere trabajar. A continuación se definen cada una de las respectivas transformaciones para este Sistema de Familias de Distribuciones de Johnson, tomadas de Chou, Polansky & Mason (1998).

3.1. Transformaciones y condiciones para la familia S_B de Johnson

Cuando los datos pertenecen a la familia S_B la transformación aplicada es:

$$Z = \gamma + \eta \ln \left(\frac{X - \epsilon}{\lambda + \epsilon - X} \right) \quad (4)$$

sujeta a:

Condiciones de los parámetros: $\eta, \lambda > 0, -\infty < \gamma < \infty, -\infty < \epsilon < \infty$.

Condiciones de la variable X : $\epsilon < X < \epsilon + \lambda$.

La estimación de los parámetros $\gamma, \eta, \epsilon, \lambda$ se encuentra en el apéndice 1.

3.2. Transformaciones y condiciones para la familia S_L de Johnson

Cuando los datos pertenecen a la familia S_L la transformación aplicada es:

$$Z = \gamma + \eta \ln(X - \epsilon) \quad (5)$$

Sujeta a:

Condiciones de los parámetros: $\eta > 0, -\infty < \gamma < \infty, -\infty < \epsilon < \infty$.

Condiciones de la variable X : $X > \epsilon$.

La estimación de los parámetros γ, η, ϵ se encuentra en el apéndice 1.

3.3. Transformaciones y condiciones para la familia S_U de Johnson

Cuando los datos pertenecen a la familia S_U la transformación aplicada es:

$$Z = \gamma + \eta \operatorname{senh}^{-1} \left(\frac{X - \epsilon}{\lambda} \right) \quad (6)$$

Sujeta a:

Condiciones de los parámetros: $\eta, \lambda > 0, -\infty < \gamma < \infty, -\infty < \epsilon < \infty$

Condiciones de la variable X : $-\infty < X < \infty$

La estimación de los parámetros $\gamma, \eta, \epsilon, \lambda$ se encuentra en el apéndice 1.

3.4. Selección de la familia a través del método de los percentiles

Se han propuesto diversos métodos para la estimación de los parámetros para este grupo de familias. El de mayor aceptación ha sido el método de los percentiles, ya que las fórmulas de estimación producidas por el mismo, son simples y explícitas.

En 1964 Johnson propone un método de estimación para los parámetros de estas familias basado en el cálculo de cuatro percentiles. Sin embargo, la selección de estos cuatro percentiles puede ser arbitraria. Autores como Bukac (1972), Bowman & Shenton (1988, 1989) citados en Chou, Polansky & Mason (1998) sugieren escoger 4 desviaciones normales estándar, $-sz, -z, z, sz$, donde $s \geq 1$ y z son constantes positivas arbitrarias.

Sea X una variable aleatoria la cual sigue una distribución miembro de alguna de las familias de Johnson (S_U, S_B, S_L). Sean q_1, q_2, q_3, q_4 , áreas bajo la curva normal estándar de $-sz, -z, z, sz$ respectivamente, por lo tanto

$$q_1 = \phi(-sz), q_2 = \phi(-z), q_3 = \phi(z), q_4 = \phi(sz),$$

donde $\phi(\cdot)$ es la función de distribución de una variable aleatoria normal estándar.

Sea x_i el q_i -ésimo cuantil de la distribución de X . Se define la razón de cuantiles QR como:

$$QR = \frac{(x_4 - x_3)(x_2 - x_1)}{(x_3 - x_2)^2} \quad (7)$$

Haciendo uso de esta razón de cuantiles se presenta el siguiente criterio para escoger entre la distribución S_B o S_U , Chou, Polansky, Mason (1998):

- Para una distribución S_B , $QR < \frac{(s-1)^2}{4}$, si $1 < s \leq 3$
- Para una distribución S_U , $QR > \frac{(s-1)^2}{4}$, si $s \geq 3$

Sobre este concepto se desarrollaron las estimaciones de los parámetros para cada una de las familias del Sistema de Familias de Distribuciones de Johnson.

En este caso se utilizan los percentiles muestrales en lugar de los poblacionales (ecuación(14)). Como ya se había mencionado, la elección de s es arbitraria, sin embargo para algún $s > 1$ a excepción de $s = 3$ la anterior regla no permite distinguir entre las tres familias, por lo tanto si se desea distinguirlas se sugiere escoger $s = 3$. Cuando esto sucede el criterio de Bowman y Shenton se reduce al criterio de Slifker y Shapiro (1980), mostrando que QR es un cociente que discrimina entre las tres familias como sigue:

- Si X tiene distribución S_B , entonces $QR < 1$.
- Si X tiene distribución S_L , entonces $QR = 1$.
- Si X tiene distribución S_U , entonces $QR > 1$.

En SPC, cuando se tiene un conjunto de datos del cual se desconoce la distribución, la cantidad QR es también desconocida, por lo tanto el criterio anterior no se puede aplicar directamente, pero para una muestra es posible estimar QR por el correspondiente estimador de la razón de cuantiles denotado como \widehat{QR} y definido de la siguiente forma:

$$\widehat{QR} = \frac{(\widehat{x}_4 - \widehat{x}_3)(\widehat{x}_2 - \widehat{x}_1)}{(\widehat{x}_3 - \widehat{x}_2)^2} \quad (8)$$

donde \widehat{x}_i es el q_i -ésimo cuantil muestral de una muestra de tamaño n . Usualmente \widehat{x}_i es la j -ésima observación ordenada, donde $j = [nq_i] + 1$ o $j = [(n+1)q_i]$, y $[.]$ representa la función mayor entero menor que el argumento.

Slifker & Shapiro (1980) sugieren el uso de $j = nq_i + 0,5$, e interpolan cuando j no es entero. Note que la función \widehat{QR} es una función determinada por las estadísticas de orden. Como \widehat{x}_i depende de z , el valor \widehat{QR} es una función que depende de z y n , entonces el criterio de Slifker y Shapiro puede ser usado reemplazando QR por su estimador \widehat{QR} , y de esta manera se puede llegar a escoger la familia de distribuciones adecuada para ajustar los datos. En 1989 un estudio de simulación realizado por Shayib muestra que cuando se conoce la distribución de los datos, el criterio de Slifker y Shapiro es válido para todos los cambios de z . En el caso muestral la variabilidad de los cuantiles muestrales evidencia que el criterio no discrimina entre las familias sin tener un error. Esto sugiere que para asegurar un buen ajuste para la muestra, se debe buscar el valor de z que proporcione el mejor ajuste a las distribuciones de Johnson y por tanto la mejor transformación de los datos, Chou, Polansky & Mason (1998).

3.5. Determinación del mejor ajuste

Existen diferentes pruebas que permiten determinar si un conjunto de datos sigue determinada distribución, ellas son llamadas pruebas de bondad de ajuste. Para el caso del Sistema de Familias de Distribuciones de Johnson la estadística utilizada por Chou, Polansky & Mason (1998) para probar normalidad es la de Shapiro Wilk (W), Conover (1980).

La selección de una familia específica de Johnson depende de la elección de z . El valor óptimo para z está definido como el valor que maximice W . El método utilizado para encontrar el z óptimo se conoce como “grid-search”. Este procedimiento requiere la especificación de un posible rango de valores óptimos de z . Aplicando este método, Chou, Polansky & Mason (1998) determinaron mediante estudios de simulación un posible rango de valores óptimos entre 0,25 y 1,25 con incremento de 0,01 es decir $S = \{z : z = 0,25; 0,26; 0,27; \dots; 1,25\}$, abarcando 101 valores para z . Este estudio sugirió no usar valores por fuera de este rango ya que producen valores- p menores a 0,05, indicando un pobre ajuste. En la práctica se tiene un conjunto de datos X_1, X_2, \dots, X_n que provienen de una distribución desconocida. Los pasos que se deben seguir para lograr el mejor ajuste son:

1. Calcular W y su correspondiente valor- p para los datos, con el fin de determinar si son normales o no. Si la normalidad es aceptada no se requiere transformación alguna. De lo contrario se sigue con el paso 2.
2. Para cada z en $S = \{z : z = 0,25; 0,26; 0,27; \dots; 1,25\}$ se calcula \widehat{QR} . Si $\widehat{QR} < 1$, se pueden ajustar las familia S_L o S_B , en otro caso S_L o S_U . Para poder realizar el ajuste se deben estimar los parámetros de acuerdo a la familia escogida y realizar la transformación.
3. Si los datos se ajustan a alguna de las familias S_U , S_L o S_B , se verifican las condiciones sobre los parámetros y el rango de la variable X . Si alguna de estas condiciones no se cumple, entonces la familia ajustada bajo ese valor z no es adecuada.
4. Para cada z en S y su correspondiente familia ajustada se calcula W y su correspondiente valor- p .
5. El mejor ajuste está asociado a la transformación con el valor W más alto. Bajo esta familia los datos transformados son aproximadamente normales estándar, permitiendo la aplicación de las técnicas SPC tradicionales, Chou, Polansky & Mason (1998).

El desarrollo metodológico del presente trabajo llevó a la realización de un programa en el paquete S-PLUS versión 2000, el cual determina el mejor ajuste de un conjunto de datos a una de las Familias del Sistema de Familias de Distribuciones de Johnson siguiendo la metodología de Chou, Polansky & Mason (1998) y transforma los datos, este módulo de programación será enviado bajo solicitud del interesado.

4. Aplicación en la industria

En una industria tan competitiva como la de las geomembranas o películas de PVC (Cloruro de Polivinilo), llevar la delantera en calidad es siempre lo deseado. Por esta razón la introducción de técnicas de Control Estadístico de Procesos es de gran importancia y acogida en este sector. Aunque cada día es más difícil tener acceso a los procesos de manufactura en la industria, es en el área de Control de Calidad donde es muy importante realizar la aplicación de las técnicas estudiadas con el fin de evaluar su alcance en casos “reales”. La aplicación de las metodologías planteadas en este trabajo se realizó sobre un conjunto de datos provenientes de un proceso de manufactura de películas a base de resinas de PVC de la compañía Filmtex S.A., específicamente sobre el brillo de su producto Madefilm WT03 siendo esta una de las características de calidad más importantes del mismo. FILMTEX S.A. es una compañía colombiana dedicada a la producción de materias primas derivadas en la forma de películas rígidas y flexibles obtenidas del PVC mediante el proceso de calandrado, para la elaboración de empaques, juguetería, muebles, protección, publicidad exterior e interior y artes gráficas entre otros.

4.1. Característica de calidad

El Madefilm WT03 es una película termoformable a base de resinas de PVC con una formulación específica que lo hace rígido o flexible, diseñado para la elaboración de puertas y gavetas de cocina, baños y bibliotecas entre otros. Es uno de los productos estrella de la compañía ya que ocupa el tercer renglón en kilos producidos por año por esta razón es importante realizar controles sobre el mismo.

Una de las características más importantes del Madefilm WT03 es el brillo del mismo, ya que por tener características de opacidad garantiza un óptimo acabado y de esta forma competitividad frente a otros productos similares. Por esta razón la característica de calidad sobre la que se monitorea el proceso es

brillo, aunque cabe aclarar que existen otras como calibre, elongación y peso. El conjunto de datos sobre el cual se aplican las cartas de control consta de 87 mediciones sobre diferentes estampados de Madefilm WT03 (granito, peral, cedro especial entre otros) que conservan las mismas características físicas excepto en estampado. Estos datos fueron recopilados durante las producciones del año 2001. El brillo de este producto tiene unas especificaciones de producción, es decir un rango en el cual debe permanecer para ser aceptado. Este rango es de 4 a 8 por ciento de luz reflejada por la lámina o brillo.

De acuerdo al manual de procedimientos de Filmtex certificada ISO 9001 y con base en la experiencia de muchos años de los departamentos técnicos y de producción se ha determinado la frecuencia de muestreo. Cada 4 horas se toma una muestra de aproximadamente 1 metro de largo y sobre esta se evalúan características como el calibre, peso, encogimiento, apariencia y brillo, registrándose un solo valor para cada una de las características, lo que sugiere que la carta de control X es la herramienta apropiada para este tipo de proceso. La medición sobre esta última se realiza a través de un brillómetro que es un aparato que emite un rayo luminoso sobre la película y calcula el porcentaje de luz reflejado bajo inclinaciones de 20, 60 y 85 grados así como combinaciones de ellos. Por lo general y para este material se trabaja brillo a 60 grados y se mide en tres puntos diferentes de la muestra: centro y los dos extremos. La medición final registrada por los operarios es el promedio de estas tres mediciones.

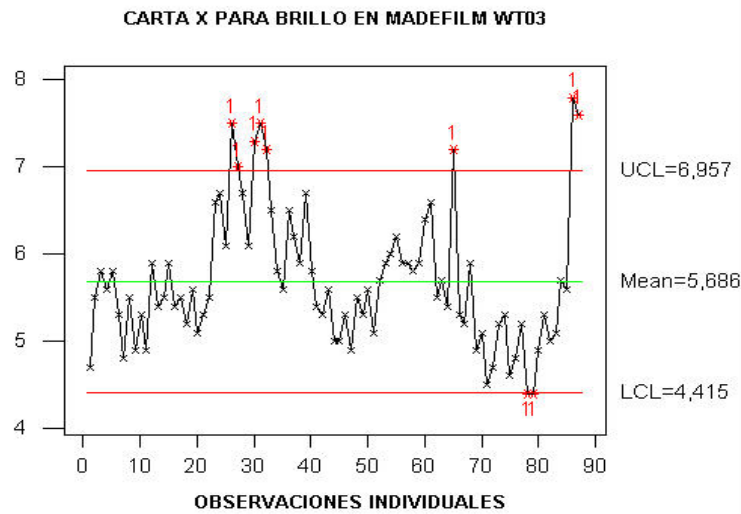
Sobre esta característica se aplicará la carta de control X , primero ignorando la distribución de los datos, posteriormente acudiendo al Sistema de Familias de Distribuciones de Johnson si no se cumple la normalidad con el fin de evaluar su funcionamiento en un caso real.

4.2. Cartas de control

El brillo sobre Madefilm WT03 es la característica de calidad bajo estudio. Se trabajó con un conjunto de 87 mediciones recolectados en el año 2001. Estos datos hacen parte de la etapa retrospectiva de la carta. El primer paso fue construir la carta de control X de la manera usual (Figura 1) sin tener en cuenta la distribución de la característica de calidad. Se calcularon los límites 3σ con $d_2 = 1,128$ es decir rangos móviles de tamaño 2.

Posteriormente se aplicó la prueba de normalidad de Shapiro Wilk sobre los datos, donde se obtuvo $W = 0,925885$ y valor- $p = 0,0000431$ indicando la no

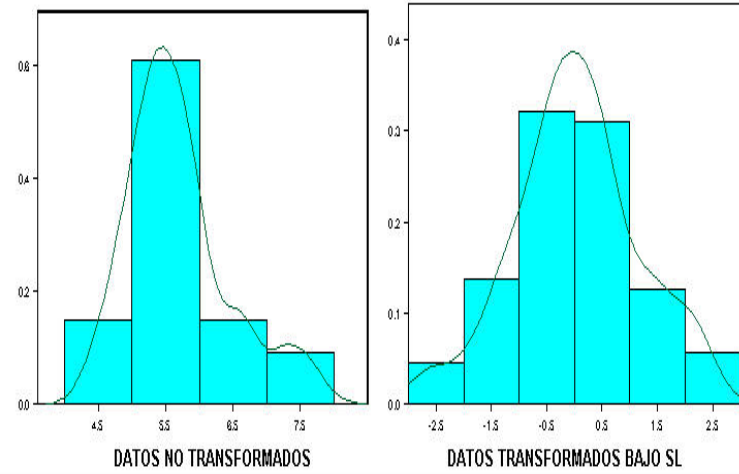
Figura 1: Carta X sobre los datos originales



normalidad de los datos. Una vez rechazada la hipótesis de normalidad se acudió al Sistema de Familias de Distribuciones de Johnson con el fin de encontrar la transformación que normalizara los datos. Se realizaron en total 139 ajustes y se determinó que el mejor era aquel asociado al valor $z = 0,76$ bajo la familia S_L . El resultado de la prueba de normalidad sobre los datos transformados bajo estas condiciones fue $W = 0,973607$ y valor- $p = 0,287115$. En la Figura 2 se presentan los histogramas para los datos de Brillo sin transformación y transformados bajo la familia S_L . Una vez normalizados los datos se construyó la carta de control X sobre ellos (Figura 3).

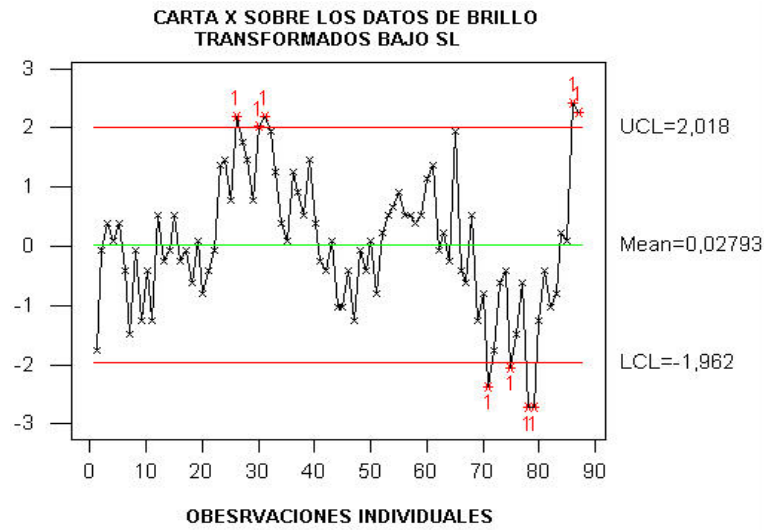
Al comparar las dos cartas X , una sobre los datos originales y otra sobre los datos transformados, es claro que guardan el mismo patrón y de igual forma muestran el punto 86 como una señal fuera de control, la cual es debida a un cambio de materiales durante el proceso según los registros de laboratorio de la planta. Con respecto al Límite de Control Superior (LCS), la primera carta muestra a los puntos 26, 27, 30, 31, 32, 65, 86 y 87 por encima él, mientras que la segunda muestra a los puntos 26, 30, 31, 86 y 87. Al revisar estos puntos dentro de los registros de laboratorio se encontró que realmente los puntos que determinaban una alteración sobre el proceso eran aquellos que revelaba la carta X sobre los datos transformados (Figura 3), producidos por cambios de materiales

Figura 2: Histograma de Brillo antes de la transformación y después de ella



y formulación del producto. De esta manera es evidente que la aplicación de la carta X sobre los datos originales (Figura 1) está incrementando el riesgo tipo I.

Con relación al Límite de Control Inferior (LCI), la carta X sobre los datos originales (Figura 1) muestra a los puntos 71, 75, 78 y 79 muy cerca del LCI , pero sólo la carta X sobre los datos transformados (Figura 3) indica que estos puntos están por debajo de él. Al revisarlos dentro del proceso se encontró que en realidad se cometieron errores de medición debido a la calibración del equipo y a un cambio de operarios. De esta manera, la carta de control X sobre los datos transformados detectó aquellos puntos en los cuales el proceso sufrió alteraciones, mientras que la carta X sobre los datos originales no, por lo tanto la aplicación de esta última está incrementando adicionalmente el riesgo tipo II. A través de este ejemplo de aplicación se evidencian las bondades de la normalización producto del Sistema de Familias de Distribuciones de Johnson, pues garantizando la normalidad de la característica de calidad, la aplicación de la teoría de las cartas de control es correcta y adecuada.

Figura 3: Carta X sobre los datos transformados bajo S_L 

5. Conclusiones

La aplicación de la carta de control X no es adecuada para un conjunto de datos cuya distribución no es normal ya que genera un incremento del Riesgo tipo I o el Riesgo tipo II, lo cuál significa un incremento en costos de producción dentro de un proceso industrial. El Sistema de Familias de Distribuciones de Johnson es una buena alternativa para el manejo de este tipo de datos ya que permite encontrar fácilmente una transformación que permita normalizarlos y bajo este nuevo conjunto de datos transformado, la aplicación de técnicas de Control Estadístico de Procesos es adecuada y correcta, sin embargo cabe aclarar que estas transformaciones son útiles no solo en el área de Control Estadístico de Procesos sino en aquellos casos en los que se requiera realizar una transformación que normalice un conjunto de datos.

6. Apéndice 1. Estimación de parámetros

6.1. Estimación de parámetros de la familia S_U de Johnson

Sean: $m = (x_4 - x_3)$, $n = (x_2 - x_1)$, y $p = (x_3 - x_2)$. La razón de cuantiles se define como:

$$QR = \frac{mn}{p^2} \quad (9)$$

Los parámetros que se deben estimar para esta familia dependen del QR en términos de m/p y n/p y son según la transformación η , γ , λ , y ϵ . Por lo tanto los parámetros están dados por Slifker & Shapiro (1980):

$$\eta = \frac{2z}{\cosh^{-1}\left(\frac{1}{2}\left(\frac{m}{p} + \frac{n}{p}\right)\right)} \quad (10)$$

donde $\eta > 0$.

$$\gamma = \eta \sinh^{-1}\left(\frac{\left(\frac{n}{p} - \frac{m}{p}\right)}{2\left(\frac{mn}{p^2} - 1\right)^{1/2}}\right) \quad (11)$$

$$\lambda = \frac{2p\left(\frac{mn}{p^2} - 1\right)^{1/2}}{\left(\frac{m}{p} + \frac{n}{p} - 2\right)\left(\frac{m}{p} + \frac{n}{p} + 2\right)^{1/2}} \quad (12)$$

con $\lambda > 0$. Finalmente

$$\epsilon = \frac{(x_z + x_{-z})}{2} + \frac{p\left(\frac{n}{p} - \frac{m}{p}\right)}{2\left(\frac{m}{p} + \frac{n}{p} - 2\right)} \quad (13)$$

6.2. Estimación de parámetros de la familia S_B de Johnson

Los parámetros que se deben estimar para esta familia dependen del QR en términos de p/m y p/n y son según la transformación η , γ , λ y ϵ . Por lo tanto los parámetros están dados por Slifker & Shapiro (1980):

$$\eta = \frac{z}{\cosh^{-1}\left(\frac{1}{2}\left[\left(1 + \frac{p}{m}\right)\left(1 + \frac{p}{n}\right)\right]^{1/2}\right)} \quad \eta > 0, \quad (14)$$

$$\gamma = \eta \operatorname{senh}^{-1} \left(\frac{\left(\frac{p}{n} - \frac{p}{m} \right) \left(\left(1 + \frac{p}{m} \right) \left(1 + \frac{p}{n} \right) - 4 \right)^{1/2}}{2 \left(\frac{p^2}{mn} - 1 \right)} \right) \quad (15)$$

$$\lambda = \frac{p \left(\left(\left(1 + \frac{p}{m} \right) \left(1 + \frac{p}{n} \right) - 2 \right)^2 - 4 \right)^{1/2}}{\left(\frac{p^2}{mn} - 1 \right)}, \quad (16)$$

con $\lambda > 0$. Finalmente

$$\epsilon = \frac{(x_z + x_{-z})}{2} - \frac{\lambda}{2} + \frac{p \left(\frac{p}{n} - \frac{p}{m} \right)}{2 \left(\frac{p^2}{mn} - 1 \right)}. \quad (17)$$

6.3. Estimación de parámetros de la familia S_L de Johnson

Los parámetros que se deben estimar para esta familia dependen del QR en términos de p , n y m y son según la transformación η , γ , y ϵ . Por lo tanto los parámetros están dados por Slifker & Shapiro (1980):

$$\eta = \frac{2z}{\ln\left(\frac{m}{p}\right)} \quad (18)$$

donde $\eta > 0$,

$$\gamma = \eta \ln \left[\frac{\frac{m}{p} - 1}{p \left(\frac{m}{p} \right)^{1/2}} \right] \quad (19)$$

$$\epsilon = \frac{(x_z + x_{-z})}{2} - \frac{p \left(\frac{m}{p} + 1 \right)}{2 \left(\frac{m}{p} - 1 \right)} \quad (20)$$

Bibliografía

Chou, Y. M., Polansky, A. M. & Mason, R. L. (n.d.), 'Transforming non-normal data to normality in statistical process control', *Journal of Quality Technology* **30**.

- Conover, W. J. (1980), *Practical Nonparametric Statistics*, 2nd edn, Grupo Editorial John Wiley and Sons.
- Lagos, I. J. (2002), *Cartas de control para datos individuales en procesos con observaciones no-normales*, Universidad Nacional de Colombia, Bogotá.
- Mittag, H. J. & Rinne, H. (1993), *Statistical Methods of Quality Assurance*, Grupo editorial Chapman and Hall, Alemania.
- Montgomery, D. C. (2001), *Introduction to Statistical Quality Control*, John Wiley and Sons, New York.
- Shore, H. (2000), 'Three approaches to analyze quality data originating in non-normal populations', *Quality Engineering* **13**(2), 377–291.
- Slifker, J. F. & Shapiro, S. S. (1980), 'The johnson system, selection and parameter estimation', *Technometrics* **32**(2), 239–246.
- Vargas, J. A. (2001), *Introducción al control estadístico de la calidad*, Universidad Nacional de Colombia, Departamento de Matemáticas y estadística, Bogotá.