ERRORES AL SELECCIONAR VARIABLES CAUSANTES DE SEÑAL DE FUERA DE CONTROL EN CARTAS DE CONTROL DE CALIDAD MULTIVARIADAS

Carlos E. Gómez
Universidad Autónoma de Colombia
Bogotá, Colombia

Las cartas de control estadístico son herramientas fundamentales en el estudio de los procesos de control de calidad a nivel de producción. Los estudiantes de ingeniería, deben tener la competencia necesaria para interpretar cartas de control, en especial cuando el proceso de control de calidad involucra muchas variables, como sucede a nivel empresarial en los procesos de producción. Se propone una estrategia que permita al estudiante adquirir las herramientas necesarias para interpretar los resultados al aplicar cartas de control a variables de control.

INTRODUCCIÓN

Las cartas de control son herramientas estadísticas utilizadas para monitorear y controlar un proceso de manufactura. Fueron creadas por Walter A. Shewhart (1924) y se emplean para mantener estándares de calidad en procesos de fabricación. Una carta de control estadístico es el procedimiento inferencial con el cual se decide si una desviación observada de la norma deseada se debe sólo al azar o a alguna causa atribuible.

Si la decisión es que la variación es aleatoria, entonces se dice que el proceso de interés se encuentra bajo control (BC). De otro modo, se juzga como fuera de control (FC) y en este caso lo que se hace, en forma general, es detener el proceso y llevar a cabo todos los esfuerzos necesarios para detectar la causa del problema.

Dado que la inferencia se basa en la probabilidad, es posible que un proceso se juzgue fuera de control cuando, de hecho, se encuentra bajo control o viceversa.
Las consecuencias de estos errores pueden ser severas; por ejemplo si se declara a un proceso como fuera de control, cuando en realidad está bajo control, se tratará de determinar una causa inexistente. Por otro lado, si el proceso en realidad está fuera de control y se permite que este continúe, el estándar de calidad deseado no se alcanzará. Se debe notar que estos errores son idénticos a los errores tipo I y tipo II conocidos en pruebas de hipótesis.

Usualmente, la determinación de una carta de control depende de la toma periódica de muestras aleatorias de tamaño n del proceso de interés, con lo que se obtiene, para cada una de éstas, un valor de alguna estadística de importancia como la media, el rango o la desviación estándar maestral. Si sólo interviene una variable en el proceso, se está tratando con cartas de control univariadas; por el contrario, si dos ó más variables intervienen en dicho proceso, se construyen cartas de control multivariadas.

Estas cartas son fáciles para construir, visualizar e interpretar, y lo más importante, tienen demostrada efectividad en la práctica. Sin embargo, con mucha frecuencia suelen ser utilizadas cuando corresponden a observaciones univariadas y se cumpla el supuesto de normalidad de la variable en proceso, lo cual no siempre es cierto.

Las cartas de control son una representación gráfica de los valores de la estadística observada, frente al número de la muestra o el período durante el cual se obtuvo ésta. Una carta de control está conformada por un límite de control superior (LCS) y un límite de control inferior (LCI), que constituyen los criterios de decisión para el proceso, es decir, una gráfica de estas, advierte que el proceso está bajo control cuando los valores de la estadística se encuentren dentro de estos límites. Si un valor de la estadística se encuentra fuera de los límites de control, se considera que el proceso está fuera de control. También se encuentra una línea central (LC) que define la norma prescrita para el proceso. Una carta de control se muestra en la siguiente figura:

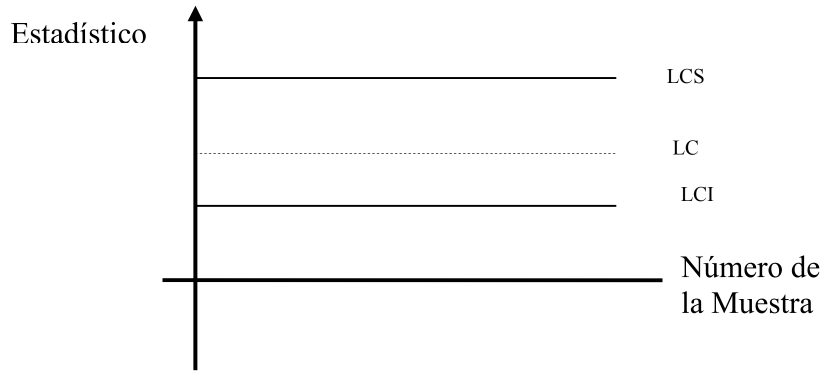

Figura 1. Prototipo de una carta de control univariada

Con una carta de control univariada se puede decidir: cuáles deben ser los valores de los límites de control, cuántas veces es necesario muestrear, es decir la frecuencia del muestreo. El tamaño de muestra está determinado por la rata de producción, y variará de período a período. Tanto el tamaño de la muestra como la frecuencia del muestreo deben ser determinados de manera conjunta con los directivos de la empresa. Una muestra grande producirá límites que quedarán más cercanos a la línea media de la carta, pues las desviaciones estándar de estadísticas como $\overline{X}$ varían inversamente con $\sqrt{n}$; es decir, entre más grande es n, más pequeña es la desviación estándar y tanto más cercanos los límites a la línea central de la carta. Además con una carta de control univariada se puede decidir que acción realizar una vez que se juzga al proceso como fuera de control.

Sin embargo, existen algunos principios generales, Shewhart argumentaba que podía alcanzarse un balance apropiado entre el costo del muestreo y la exactitud del estimador, si las muestras tienen un tamaño de cuatro o cinco observaciones cada vez. Igualmente, los límites de control tres-sigma han demostrado ser muy satisfactorios, estos límites se explican de la siguiente manera:

Si se asume que la distribución del estadístico, digamos X , es normal, se acostumbra a usar estos límites de control tres-sigma (3σ) , esto es, a partir del valor prescrito para el proceso (línea central), se suma y se resta 3 desviaciones estándar de X para obtener los límites superior e inferior respectivos de la carta.

La carta de control univariada de mayor interés para el desarrollo del presente trabajo es la carta de control para la media, esta se conoce como carta $\overline{X}$.

Pero, en la práctica se consideran de manera conjunta mas medidas de calidad multivariadas que univariadas, ya que generalmente la calidad de un producto está determinada por más de una característica de calidad, es decir, la calidad generalmente es medida por el control simultáneo de varias variables aleatorias posiblemente correlacionadas, esto debido a que hay situaciones en las cuales es necesario controlar dos o más características de un producto o proceso. Por ejemplo, la calidad de un cierto tipo de pastillas farmacéuticas está determinada por su peso, grado de dureza, espesor, anchura y longitud. Estas características están correlacionadas, ya que existe asociación entre las variables, por ejemplo el peso depende del grado de dureza, etc.
Si se desea detectar cambios, por ejemplo, en cada media del proceso a diferentes intervalos de tiempo, las cartas de control univariadas no son las adecuadas.

Luego, se deben tener cartas de control que no solo midan directamente características de calidad multivariadas, si no que incorporen la información presente en los datos, como es el caso de la covariabilidad. Entre las cartas de control multivariadas está la carta de control T^2 de Hotelling que se utiliza para detectar señales fuera de control en vectores de medias, o las cartas a partir de componentes principales y otras.

El acercamiento multivariado al control de calidad fue publicado por primera vez en 1947 y 1951 por Hotelling, quien propuso el uso de cartas de control multivariadas basadas en la estadística T^2 suponiendo que la distribución asociada a las variables aleatorias es normal multivariada.

En este trabajo se presenta una metodología para detectar las variables causantes de la señal de fuera de control en un proceso de control de calidad multivariado y los frecuentes errores que se cometen al intentar utilizar cartas de control univariadas de manera independiente para cada variable involucrada en el proceso.

GENERALIDADES

Alt(1985) y Alt y Smith(1988) establecieron dos fases para el proceso de construcción de una carta de control multivariada para la media. La Fase I, a su vez se divide en dos etapas. En la etapa 1 o etapa retrospectiva se prueba si el proceso estaba bajo control cuando se recogieron los primeros subgrupos. En la etapa 2 o etapa prospectiva se prueba si el proceso permanece bajo control cuando futuros subgrupos sean seleccionados. En la Fase II se usa la carta de control para detectar desvíos del proceso respecto de su valor estándar $\vec{\mu}_0$.

Se trabaja con la “fase II” del problema de control de calidad, que utiliza cartas de control especialmente para detectar cualquier salida del proceso p-dimensional de los valores estándar conocidos ($\mu_0;\Sigma$).

Las ideas desarrolladas son, no obstante, fácilmente extendibles para el caso donde ($\mu_0;\Sigma$) puedan ser estimadas. Además, se supone, que el proceso p-dimensional está distribuido normalmente y que los datos disponibles son medias muestrales, $\overline{X}'s$, basadas en muestras de tamaño n, seleccionadas del proceso.

Luego, la población en control, notada por $\prod_0$, está distribuida normalmente, así $\prod_0 \approx N_p(\mu_0,\Sigma)$ y la familia de poblaciones fuera de control, notada por $\prod$, está distribuida normalmente, es decir $\prod \approx N_p(\mu;\Sigma)$, donde, $\mu \neq \mu_0$. El procedimiento hace monitoreos solo para algún cambio en la media μ_0.

La decisión sobre si el proceso está bajo control o no, se obtiene mediante la distancia cuadrada de Hotelling T^2, la cual está dada por:

$$T^2(\bar{x}) = n(\mu_0 - \bar{x})^t \ \Sigma^{-1}(\mu_0 - \bar{x})$$

que es equivalente a verificar la hipótesis: $H_0 : \mu = \mu_0$ v.s $H_1 : \mu \neq \mu_0$.

Con $\bar{x} \in \prod_0$, $T^2 \approx \chi_p{}^2$ y con $\bar{x} \notin \prod_0$, $T^2 \approx \chi^2(p,\lambda)$, donde λ es el parámetro de no centralidad de la distribución ji-cuadrado con p grados de libertad y está dado por:

$$\lambda = n(\mu_0 - \mu)^t \, \Sigma^{-1}(\mu_0 - \mu).$$

Usando la *distribución central* se selecciona un corte (límite) K correspondiente a cualquier percentil deseado α de $\chi_p{}^2$ de manera que la regla de decisión es:

Si	$T^2 \leq K$;	**CONTINÚE**,	proceso en control(BC)
Si	$T^2 > K$;	**PARE**,	proceso fuera de control(FC)

Se mide la efectividad del procedimiento utilizando la distribución de la longitud de racha, y, específicamente, de la longitud promedio de racha(ARL).

Para cualquier carta Shewhart, la ARL es: $ARL = 1/p$

donde p es la probabilidad de que un punto caiga fuera de los límites de control. Así para la carta de $\overline{X}$ con los límites 3σ acostumbrados:

$$p = 1 - 0.9973 = 0.0027$$

es la probabilidad de que un punto se sitúe fuera de los límites cuando el proceso está bajo control. Donde $0.9973 = p(-3 < Z < 3)$

Entonces, la ARL de la gráfica de $\overline{X}$ es: $ARL = 1/p = 1/0.0027 = 370$,

cuando el proceso está bajo control. Es decir, aun cuando esto último suceda, se generará una señal de bajo control cada 370 muestras, en promedio.

Para el procedimiento T^2, la distribución de la longitud de racha es una distribución geométrica con parámetro P, donde P es la probabilidad de no rechazo de que el proceso está bajo control, (BC) con:

$$ARL_{T^2}(\mu) = \frac{1}{(1-P)} = 1/\Pr(parar).$$

Luego:

$$P = \Pr[T^2 \leq K / \bar{x} \in N_P(\mu; \Sigma/n)] = \Pr[\chi^2(\lambda, p) \leq K / \lambda]$$

y

$$ARL_{T^2}(\mu) = 1/\Pr[\chi^2(\lambda, p) > K]$$

En particular:

$$ARL_{T^2}(\mu_0) = 1/\Pr(\chi_p{}^2 > K) = \frac{1}{\alpha}$$

USO DE LA CARTA T^2

La ventaja de utilizar el estadístico T^2 es su propia reflexión sobre la estructura de correlación de las poblaciones.(Están implícitas las correlaciones de las poblaciones, ya que Σ es conocida) aunque también se presenta el caso, con un buen estimador de Σ. En la práctica, el no tener en cuenta la correlación entre las variables es la desventaja principal al usar cartas independientes para monitorear cada una de las p características de calidad de manera individual. Aún haciendo pruebas simultáneas sobre cartas individuales para permitir que cada error de tipo I sea α/p, en lugar de α, es una dificultad.

Se considera el caso p = 2, $\alpha = 5\%$, Con la carta T^2 y para todo ρ, que denota la correlación entre las variables:

$$ARL_{T^2}(\mu_0) = 1/\Pr(parar) = 1/\alpha = 20$$

es el número de puntos graficados en la carta hasta que aparezca una señal fuera de control.

Como ilustración, y usando dos cartas individuales (ID) y la regla para detenerse : "Si alguna observación cae por fuera de los límites en alguna de las cartas, PARE, el proceso está fuera de control" ,se tiene para valores exactos de α, sobre cada carta :

- Con $\rho = 0$; $\alpha = 5\%$; $\Pr(parar) = 2\alpha - \alpha^2 = 0.0975$

Se sabe que la probabilidad total de error tipo I es:

$$1-(1-\alpha)^2 = 1-(1-2\alpha+\alpha^2) = 1-1+2\alpha-\alpha^2 = 2\alpha-\alpha^2 = 0.0975$$

$$ARL_{ID}(\mu_0) = 1/0.0975 \approx 10.26$$

es el número de puntos graficados, en promedio, en la carta hasta que aparezca una señal fuera de control.

Ahora, con valores $\alpha/2$ en cada carta individual, se tiene:

- Con $\rho = 0$, $\boldsymbol{\alpha = \frac{5}{2}\%}$; $ARL_{ID}(\mu_0) = 20.25$

Sin embargo, al tratar de solucionar el problema de la ARL para cartas individuales, aumenta la tendencia a clasificar incorrectamente las observaciones como se ilustra en a siguiente figura 2:

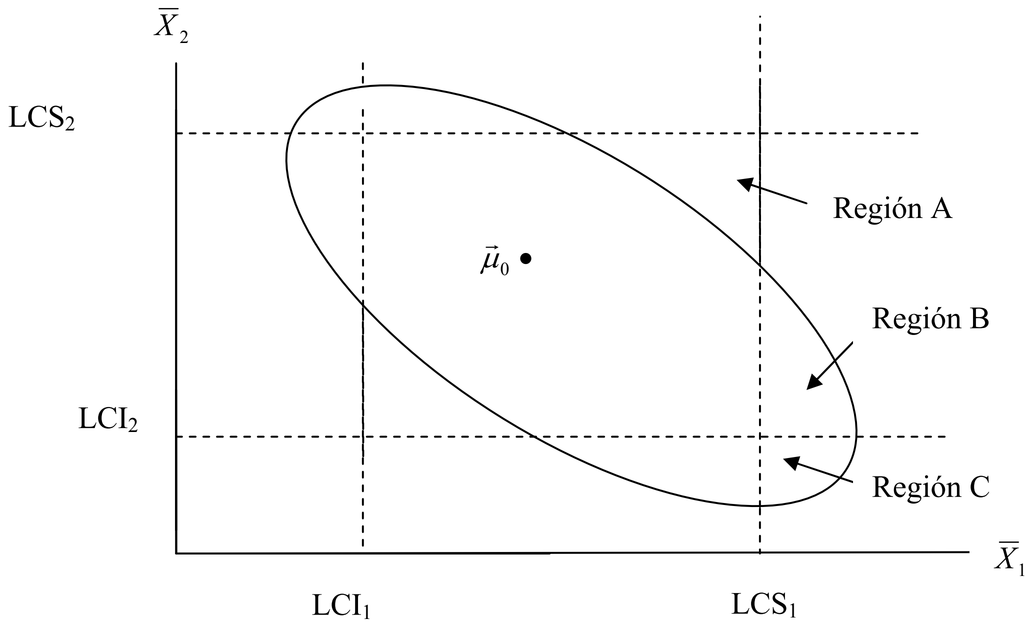

Figura 2: Regiones de Control Elíptica y Rectangular

Si el par de medias muestrales se ubica dentro de la región rectangular de intersección de las dos cartas de control, el proceso se considera bajo control, sin embargo, el uso de cartas de control por separado o el equivalente a la región rectangular en la figura 2 puede ser muy delicado para la toma de decisiones a nivel multivariado. Se observa que la región de control es realmente de naturaleza elíptica y el proceso se juzga como fuera de control solo si el par de medias $(\bar{x}_1, \bar{x}_2)$ queda por fuera de la elipse.

Al utilizar la región rectangular, sin embargo, se puede concluir erróneamente que ambos procesos de medias están bajo control, como lo muestra la figura 2 en la Región A. Esta región se podría considerar como de rechazo multivariado (de "no rechazo" univariado), de que $H_0 : \vec{\mu} = \vec{\mu}_0$ donde $\vec{\mu}_0$ es el vector de medias especificado.

La región B indica que una variable está en control, en este caso X_2, y la otra, X_1 está fuera de control.

La región C indica que ambas variables están fuera de control y se podría considerar como de rechazo univariado y de no rechazo multivariado de que $H_0 : \vec{\mu} = \vec{\mu}_0$,donde $\vec{\mu}_0$ es el vector de medias especificado.

El grado de correlación entre las dos variables afecta el tamaño de estas regiones y sus respectivos errores, más aún, la probabilidad de que ambas medias muestrales estén dentro de la región elíptica, cuando el proceso se encuentra bajo control es exactamente $1-\alpha$, mientras que la probabilidad de que estén en la región rectangular es menor que $1-\alpha$.

Se observa que la región de control real es de naturaleza elíptica y el usar límites $\alpha/2$ hace que los rectángulos sean mas grandes y esto puede conducir a conclusiones erradas.

En la figura 2, se observa como las cartas individuales clasifican incorrectamente observaciones en las regiones A, B y C.

El tamaño y tipo de estas regiones de mala clasificación, dependen del grado de correlación entre las variables y de los valores de α usados en las cartas individuales. En la práctica, se deberían graficar ambas, la carta T^2 y las p cartas individuales, con valores dados de α, aunque la decisión de control se base solamente sobre la carta T^2.

Otra razón para utilizar el procedimiento T^2 es la facilidad de cálculo y de construcción simple de una carta de control que solo requiere comparar valores T^2 con el del corte K, (de la distribución central, para α dado), como se ilustra en la figura 3:

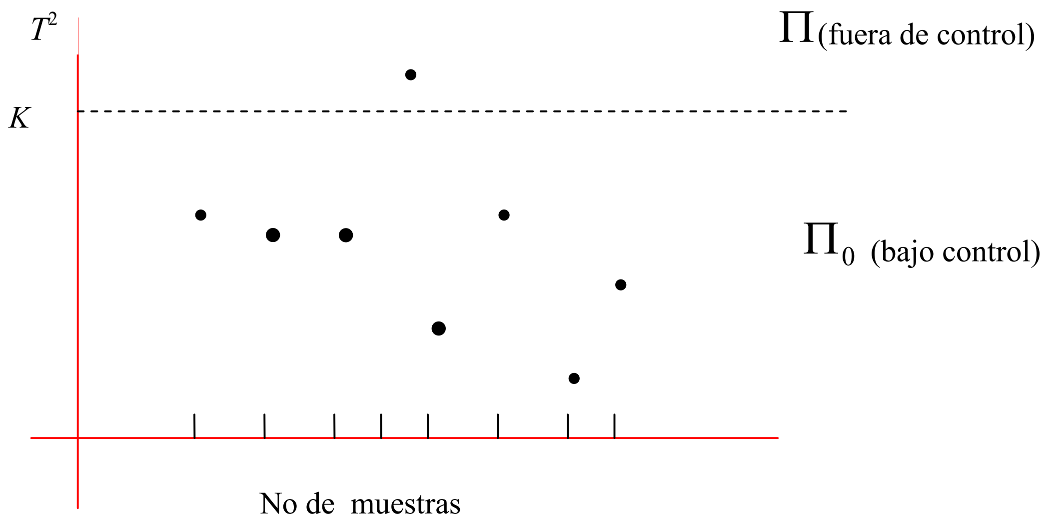

Figura 3: Carta de Control T^2 con corte K

CONCLUSIONES

Las deficientes conclusiones obtenidas de las cartas individuales muestran su ineficacia como un método de control de calidad multivariado y como un método de selección de variables fuera de control.

Sin embargo las cartas individuales (ID) pueden ser consultadas para ver la dirección de las variables identificadas como fuera de control.

REFERENCIAS

Alt, F. B. and Smith, N.D., "Multivariate Process Control," in: P.R. Krishnaiah and C. R. Rao, eds, Handbook of Statistics, vol 7, North-Holland, Amsterdan,333-351. 1988.

Anderson, T.W., "An Introduction to Multivariate Statistical Analysis", 2ed, John Wiley and Sons , New York.1984.

Diaz, L.G., "Estadística Multivariada, Inferencia y métodos" Universidad Nacional de Colombia. 2002.

Duncan, A.J., "Quality Control and Industrial Statistics", 5ed. Richard D. Irwin, Homewood, IL..1986.

Murphy,B.J., "Selecting out of Control variables with T^2 Multivariate quality Control procedures" Journal The Statistician No 36, 571-583. 1987.

Regina Y.Liu., "Control Charts for Multivariate Processes," Journal of the American Statistical Association, December 1995, vol. 90, No 432. Theory and Methods.1995.