Maria Isabel Flury
Cristina A. Barbiero
Instituto de Investigaciones Teóricas y Aplicada., Escuela de Estadística.

APLICACIÓN DE TECNICAS DE CONTROL MULTIVARIADO EN PROCESOS INDUSTRIALES

1) INTRODUCCIÓN

Las nuevas características de muchos procesos industriales y los requisitos cada vez más exigentes de los mercados, hacen que los métodos estadísticos tradicionales para la mejora de la calidad y de la productividad, resulten a veces insuficientes.

En cualquier proceso de producción, existe siempre cierto grado de variabilidad inherente o ruido de fondo, resultado del efecto acumulativo de muchas pequeñas causas, esencialmente incontrolables. En el marco del control estadístico de la calidad, esta variabilidad natural se denomina *"sistema estable de causas fortuitas"*. Un proceso que funciona sólo bajo este sistema de causas, se considera "*bajo control estadístico*".

Otro tipo de variabilidad que en general resulta mayor que el ruido de fondo y que normalmente representa un nivel inaceptable del funcionamiento del proceso, se debe a la presencia de *"causas asignables".* Un proceso que funciona en presencia de estas causas, de considera "*fuera de control estadístico".*

Uno de los objetivos más importantes del control estadístico de procesos es detectar rápidamente la ocurrencia de causas asignables , a fin de poder investigar y tomar acciones correctivas antes de la producción de gran cantidad de piezas disconformes. La meta final es reducir a un mínimo la variabilidad del proceso. El diagrama de control es una de las herramientas más efectivas para lograr esta meta.

En la actualidad, los procedimientos automáticos de inspección permiten medir con relativa facilidad varias características de calidad. Los métodos estadísticos de control de procesos multivariados son valiosos en estas circunstancias. Ellos toman en consideración la relación existente entre las variables, generando algoritmos potentes que son sensibles a la presencia de causas asignables, pobremente detectadas por diagramas de control univariados.

2) OBJETIVOS

El objetivo del trabajo es enfocar aspectos relativos a la implementación de gráficos de control multivariados basados en la estadística T^2 de Hotelling y presentar en forma breve distintos procedimientos para interpretar señales de fuera de control.

3) ETAPAS EN LA IMPLEMENTACION DEL CONTROL MULTIVARIADO DE PROCESOS

La implementación del control de calidad multivariado requiere considerar las siguientes etapas:

1) Determinación de los “sujetos de control”

En el caso multivariado, en el cual se evalúan “p” dimensiones, el vector de observaciones para cada item, puede consistir en datos recolectados en tiempos o lugares diferentes, con distintos instrumentos de medición. Por lo tanto, la elección de los “sujetos” a controlar deberá realizarse teniendo en consideración las caracteristicas del proceso a ser monitoreado, así como las del sistema de medición.

2) Identificación de las variables a medir.

3) Establecimiento de metas o valores objetivos (internos o externos) para cada una

de las variables consideradas.

Existen diversas alternativas para establecer valores objetivos para las variables monitoreadas. Pueden emplearse *valores determinados externamente*, derivados de consideraciones ajenas al proceso. Aquí, el control de calidad tiene como objetivo, adherirse a dichos estándares o satisfacer metas establecidas por la competencia. Otro procedimiento consiste en generar *valores objetivos internos*, basándose en *estudios de capacidad de proceso*. El objetivo es aquí detectar observaciones anómalas o outliers. Por último, puede utilizarse una *muestra base o de referencia* recolectada previamente. En este caso , el control de calidad determina si la observación (o grupo de observaciones) que está siendo evaluada, es consistente con la muestra de referencia.

4) Creación de un mecanismo de recolección de datos para evaluar la performance del proceso.

Esto se refiere tanto a los instrumentos de medición, como a la programación de ensayos y a la determinación de los métodos mediante los cuales se informa de los resultados, al operador encargado del proceso.

5) Evaluación de la performance del proceso

El control multivariado requiere sincronizar las actividades, de manera que la información llegue al lugar adecuado, en el tiempo correcto.

6) Interpretación de la diferencia entre la performance real y la meta u objetivo.

Cuando se observa una condición de salida de control, debe investigarse qué variables la produjeron. Para ello pueden usarse en forma simultánea gráficos uni y multivariados, el método de componentes principales y otros procedimientos, tales como el step-down.

7) Toma de acciones correctivas.

4) LA ESTADISTICA T^2 DE HOTELLING

Si se estudian “p” características de calidad que siguen una distribución normal multivariada con vector de esperanzas m y matriz de covariancias Σ, puede establecerse un diagrama de control basado en la distribución χ^2 en el caso de Σ conocida. La estadística

$\chi^2 = c\,(X-m)'\,\Sigma^{-1}(X-m)$ sigue una distribución χ^2_p .

Si la matriz Σ, es estimada por la matriz de covariancias muestral, la estadística

T^2 = c (X-m)' S^{-1}(X-m) sigue una distribución **T^2** de Hotelling con (n-1) grados de libertad, siendo n el número de observaciones con que se calcula S. Puede apreciarse que T^2 es una distancia al cuadrado, obtenida como una constante multiplicada por una forma cuadrática, donde:

c: constante que depende del tamaño de muestra o subgrupo a partir del cual fue estimada la matriz de covariancias.

X: vector de observaciones

m: vector de valores objetivos

S: matriz de covariancias muestral

En general, cuanto mayor es el valor de T^2, más distante está la observación del valor objetivo. Si las variables son independientes, la matriz de covariancias es una matriz diagonal, y la estadística se hace proporcional a la suma de cuadrados de las variables estandarizadas.

Se podrá entonces, fijar como límite superior de control, el valor de $\mathbf{T}^2_{n-1,\alpha}$, y luego graficar las observaciones que se obtienen en los diferentes momentos, entendiendo como señal de fuera de control, la presencia de un valor observado superior a $\mathbf{T}^2_{n-1,\alpha}$.

Los dos problemas principales que surgen al aplicar el gráfico T^2 son:

a) la escala que se utiliza para graficar los valores, no está relacionada con la escala de ninguna de las variables involucradas.

b) Cuando la estadística T^2 excede el límite superior de control, el usuario no sabe cuál o cuáles son las variables responsables de la señal de fuera de control. Por eso se aconseja llevar en forma paralela, gráficos de control univariados para cada característica.

5) DISTINTOS PROCEDIMIENTOS PARA LA DETERMINACION DEL VECTOR DE VALORES OBJETIVOS Y LA ESTIMACION DE LA MATRIZ DE COVARIANCIAS .

Procedimiento 1: valores objetivos asignados externamente

En este caso, el vector de valores objetivos $\mathbf{m_0}$ es fijado externamente, ya sea a través de valores nominales de especificaciones de ingeniería o a través de valores obtenidos en estudios anteriores.

- **Si los datos forman una única muestra**, la matriz de covariancias puede ser estimada a través de la siguiente expresión:

$$S = \sum_{i=1}^{n} (X_i - \overline{X})(X_i - \overline{X})' / (n-1)$$

La estadística T^2 **para cada observación** es:

$$T^2 = (X_i - m_0)' S^{-1} (X_i - m_0)$$

Para determinar si un punto cae fuera de control, el valor de T^2 debe compararse con un límite superior de control, dado por:

$$L.S.C. = \frac{p(n-1)}{n-p} \quad F_{\alpha,p,n-p}$$

Donde F es el valor de la estadística F de Snedecor con p y (n-p) grados de libertad y nivel de significación del α %.

Lo efectuado, equivale a realizar un test de hipótesis, en el cual se postula que el vector de medias observado no difiere del dado externamente. El rechazo de esa hipótesis queda manifestado por la presencia de un punto por encima del límite superior de control.

El test multivariado tiene una potencia mayor que la de los tests univariados que podrían llevarse a cabo para cada una de las p variables involucradas.

- **Si los datos se disponen en k subgrupos de tamaño n_j,** deberá usarse como estimación , una matriz de covariancias amalgamada S_p :

$$S_p = \frac{\sum_{J=1}^{K}(n_j-1)S_j}{\sum_{j=l}^{k}(n_j-1)}$$

Si los n_j son iguales:

$$S_p = \frac{\sum_{j=l}^{k} S_j}{k}$$

La estadística T^2 **para cada uno de los subgrupos** es:

$$T^2_{Mj} = n(\overline{X}-m_0)'S_p^{-1}(\overline{X}-m_0)$$

El límite superior de control es:

$$L.S.C. = \frac{pk(n-1)}{k(n-1)-p+1}F_{\alpha;p;k(n-1)-p+1}$$

En este caso, también puede calcularse una estadística T^2_{Dj} para evaluar la variabilidad dentro de cada subgrupo :

$$T^2_{Dj} = \sum_{i=l}^{n}(X_{ij}-\overline{X}_j)'S_p^{-1}(X_{ij}-\overline{X}_j)$$

El límite superior de control para esta estadística resulta:

$$L.S.C. \cong (n-1)\chi^2_{\alpha;p}$$

Donde χ^2 es el valor de la estadística chí cuadrada con p grados de libertad.

La variabilidad total T^2_{0j} puede descomponerse en la suma de las variabilidades entre subgrupos y dentro de subgrupos:

$$T^2_{0j} = T^2_{Mj} + T^2_{Dj}$$

Haciendo un paralelo con gráficos univariados, el T^2_{Mj} es el equivalente a un gráfico de promedios que controla la variabilidad entre subgrupos y el T^2_{Dj} se correponde con un gráfico de rangos o desvíos, que controla la variabilidad dentro de cada subgrupo.

Procedimiento 2: valores objetivos determinados internamente

En este caso, los valores objetivos son determinados en forma interna, a través de un estudio de capacidad de procesos multivariado.

Los estudios de capacidad de proceso incluyen:

- Determinación de los límites del proceso y de las variables que caracterizan la salida.
- Determinación del método de muestreo y formación de subgrupos racionales
- Análisis de causas y efectos, que relacionan las salidas del proceso con parámetros internos y factores de control (diagrama de Ishikawa, matrices Quality Function Deployment).
- Recolección y análisis de datos, mediante el empleo de gráficos univariados, multivariados y otros .
- Detección y eliminación de causas asignables de variación.
- Evaluación del modelo probabilístico al cual responde el modelo. Chequeo del supuesto de normalidad multivariada.
- Cálculo de índices de performance del proceso y de índices de capacidad de proceso. Los primeros se calculan con todos los datos, sin considerar la estabilidad a lo largo del tiempo. Los segundos se calculan con datos recolectados en un período corto de tiempo y requieren un proceso estable, distribución normal e independencia de las observaciones.

- **Si los datos forman una única muestra**, la matriz de covariancias se estima de igual forma que en el procedimiento anterior.

La estadística T^2 , **para cada observación** es :

$$T^2 = (X_i - \overline{\overline{X}})' S^{-1} (X_i - \overline{\overline{X}})$$

donde $\overline{\overline{X}}$ es el vector promedio para toda la muestra. Dado que en esta estadística $(X_i - \overline{\overline{X}})$ no es independiente de S, su distribución corresponde a una distribución Beta, y el límite superior de control es:

$$L.S.C = (n-1)\ \beta_{\alpha;p/2;(n-p-1)/2}$$

Wierda(1994) y Bruyns(1992) propusieron dos alternativas para construir la estadística T^2 , de tal manera que ésta se corresponda con una distribución F. Dichas estadísticas requieren calcular para cada observación la matriz de covariancias y el vector de medias .

Desde el punto de vista práctico las diferencias entre los tres métodos suele ser despreciable, a menos que el número de observaciones sea muy pequeño.

- **Si los datos se disponen en k subgrupos de tamaño n**, deberá usarse como estimación de la matriz de covariancias , la matriz S_p ya definida.

La estadística T^2 **para cada uno de los subgrupos** es:

$$T^2_{Mj} = n\,(\overline{X}_j - \overline{\overline{X}})' S_p^{-1} (\overline{X}_j - \overline{\overline{X}})$$

El límite superior de control es :

$$\text{L.S.C.} = \frac{p(k-1)(n-1)}{k(n-1)-p+1} F_{\alpha;p;kn-k-p+1}$$

De igual forma que en el procedimiento anterior, la variabilidad dentro de cada subgrupo puede calcularse como:

$$T^2_{Dj} = \sum_{i=l}^{n} (X_{ij} - \overline{X}_j)' S_p^{-1} (X_{ij} - \overline{X}_j)$$

El límite superior de control para esta estadística resulta:

$$\text{L.S.C.} \cong (n-1)\chi^2_{\alpha;p}$$

Donde χ^2 es el valor de la estadística chí cuadrada con p grados de libertad.

La variabilidad total T^2_{0j} puede descomponerse en la suma de las variabilidades entre subgrupos y dentro de subgrupos:

$$T^2_{0j} = T^2_{Mj} + T^2_{Dj}$$

Detectados y eliminados los puntos fuera de control, se recalcula el vector de medias y la matriz de covariancias. Estos datos sirven como referencia para tests futuros del proceso productivo.

Debe mencionarse una característica especial en estos casos: cuando el número de variables es muy grande en relación al número de observaciones disponibles, los tests tienden a favorecer la hipótesis nula (proceso bajo control) y a presentar menor potencia , por ello se recomienda continuar el análisis de datos sobre subconjuntos particulares de variables (físicamente relacionadas) e investigar causas de variación potenciales que puedan afectar el proceso analizado.

NOTA:

En la práctica es preferible basar el análisis en datos agrupados, ya que la estimación de la matriz de covariancias a partir de una matriz amalgamada resulta más confiable. Se calcula la matriz de covariancias para cada subgrupo racional y luego se procede a promediarlas para obtener la matriz amalgamada. Los subgrupos racionales son más adecuados para detectar cambios en el vector de medias.

Pandit (1973) y Anderson (1970) sugieren un método para estimar la matriz de covariancias cuando no se dispone de subgrupos. El mismo consiste en forzar una muestra de tamaño 2 tomando el vector de diferencias entre observaciones sucesivas. Si la muestra es de tamaño n, habrá n–1 diferencias. La matriz de covariancias del proceso se estima construyendo una nueva matriz que tenga en cuenta las diferencias entre observaciones sucesivas:

$$S=\frac{0.5\sum_{i=1}^{n-1}(X_i-\overline{X})'(X_i-\overline{X})}{n-2}$$

Esta estimación, conjuntamente con el vector de medias de las diferencias, es utilizado para construir la estadística T^2. Dicha estadística arrojará valores más grandes que los que se obtendrían tomando toda la muestra en forma conjunta, haciendo más probable la aparición de puntos fuera de control.

El límite superior de control está basado en una estadística χ^2 con p grados de libertad.

Procedimiento 3: valores objetivos determinados a partir de una muestra de referencia.

En este caso el objetivo es comparar, o determinar diferencias entre valores de una muestra recientemente recolectada, y valores objetivos determinados a partir de una **muestra base o de referencia.**

Se supone que la **muestra de referencia** está compuesta por n_2 observaciones independientes: $X_1, X_{2,\ldots,} X_{n2}$, distribuidas según una normal p variada, con vector de medias μ_2 y matriz de covariancias Σ_2.

La **muestra de ensayo** está compuesta por n_1 observaciones independientes $Y_1, Y_{2,\ldots,}$ Y_{n1}, distribuidas según una normal p variada, con vector de medias μ_1 y matriz de covariancias Σ_1.

- Para **datos sin agrupar**.

Se supone que $\Sigma_1=\Sigma_2=\Sigma$ y que $\overline{X}$ e $\overline{Y}$ son los vectores de medias correspondientes a cada muestra.

Se tendrá que :

$$(\overline{Y} - \overline{X}) \sim N_P\{\mu_1 - \mu_2; (\frac{1}{n_1} + \frac{1}{n_2})\Sigma\}$$

Si Σ **es conocido** , la estadística T^2 es:

$$T_M^2 = \frac{n_1 n_2}{n_1 + n_2}(\overline{Y} - \overline{X})'\Sigma^{-1}(\overline{Y} - \overline{X})$$

que se distribuye como una χ^2 con p grados de libertad y parámetro de no centralidad:

$$\lambda = \frac{n_1 n_2}{n_1 + n_2}(\mu_1 - \mu_2)'\Sigma^{-1}(\mu_1 - \mu_2)$$

bajo la hipótesis nula $(\mu_1 = \mu_2)$, el parámetro de no centralidad es cero.

Si Σ **es desconocido** , se podría estimar a través de una matriz S amalgamada que contenga los datos de ambas muestras. Sin embargo es muy poco frecuente combinar la información de ambas muestras, por lo que Σ suele estimarse sólo con los datos de la muestra base (S_x).

La estadística T^2 , **para cada observación** es :

$$T^2 = n_1(\overline{Y} - \overline{X})'S_X^{-1}(\overline{Y} - \overline{X})$$

El límite superior de control es:

$$L.S.C. = \frac{p(n_2 - 1)(n_1 + n_2)}{n_2(n_2 - p)} F_{\alpha; n_2 - p}$$

- Para **datos agrupados**.

Ambas muestras se hallan divididas en k subgrupos de tamaño n. La matriz de covariancias puede estimarse a través de una matriz amalgamada $S_{Xp,}$ obtenida en base a la muestra de referencia . El vector de medias de dicha muestra se simboliza $\overline{\overline{X}}$.

La estadística T^2 **para cada subgrupo** de la muestra ensayada, es:

$$T_{Mj}^2 = (\overline{Y_j} - \overline{\overline{X}})\ S_{Xp}^{-1}(\overline{Y_j} - \overline{\overline{X}})$$

Este valor debe compararse con:

$$L.S.C. = \frac{p(k+1)(n-1)}{k(n-1)-p+1} F_{\alpha;kn-k-p+1}$$

La variabilidad dentro de los subgrupos se calcula con:

$$T^2_{Dj} = \sum_{i=l}^{n} (Y_{ij} - \overline{Y_j}) S^{-1}_{Xp} (Y_{ij} - \overline{Y_j})$$

La variabilidad total se obtiene como suma de las dos estadísticas anteriores. En situaciones prácticas, puede ser de interés evaluar qué porcentaje de la variabilidad total representa cada una de ellas.

Cabe destacar que en los tres procedimientos detallados, no sólo cambia la forma de cálculo de la estadística T^2 , sino que también se modifica el numerador del límite superior de control.

6) INTERPRETACIÓN DE SEÑALES FUERA DE CONTROL.

Cuando el valor de T^2 supera el límite de control se debe iniciar una búsqueda para detectar qué variables fueron las responsables de dicha señal.

Para ello puede recurrirse a distintos procedimientos, que se mencionan brevemente a continuación. Su análisis detallado será objeto de otros trabajos.

Procedimiento step-down: exige realizar un ordenamiento previo de las variables incluidas en el análisis. Consiste en efectuar tests en forma sucesiva, contrastando hipótesis condicionadas para un subgrupo específico de variables. El condicionamiento está en suponer que los promedios de las variables de los subconjuntos previos al analizado, son iguales a los de la población de referencia.

El método da información con respecto a los subconjuntos previos a aquél en el que se detectó la falta de control, pero no la aporta con respecto a subconjuntos posteriores .Los tests pueden llevarse a cabo utilizando distintos niveles de significación para cada subgrupo.

Se sugiere en casos prácticos colocar en primer lugar aquellas variables que a priori tienen menor probabilidad de cambio. No puede utilizarse en los casos en los cuales hay una falta total de información sobre el ordenamiento a priori. Como alternativa, pueden emplearse gráficos univariados.

Procedimientos basados en la descomposición de la estadística T^2: no suponen un ordenamiento previo de las variables. La descomposición se basa en el empleo de métodos de regresión en los que una variable dada puede ser estimada en función de las restantes. En esta alternativa existen problemas con respecto a la determinación del nivel de significación total.

Procedimientos gráficos: el software actuamente disponible permite utilizar, además de las técnicas gráficas univariadas tradicionales, otras técnicas gráficas especiales para el control de calidad multivariado, tales como star plot, sun plot, scatterplot y Mp-charts.

Componentes principales: consiste en formar nuevas variables que son combinaciones lineales de las p variables involucradas. El objetivo es que unas pocas componentes principales capten la mayor parte de la variabilidad de los datos reduciendo así la dimensión del estudio. Tiene como desventaja el perder la identidad de las variables originales, aunque en muchas situaciones prácticas puede conducir a nuevas dimensiones que pueden interpretarse naturalmente.

7) CONSIDERACIONES FINALES

Los métodos tradicionales de control estadístico de procesos se desarrollaron teniendo en cuenta la posibilidad de evaluación de unas pocas características de calidad.

La disponibilidad de una gran cantidad de información on-line hace que, actualmente, los métodos tradicionales del control estadístico de procesos sean insuficientes . En estas situaciones los métodos de control multivariado pueden aprovechar toda la información disponible y permitir la construcción de modelos para el análisis de procesos y la detección precoz de fallas.

Si bien los métodos de control multivariado se adaptan mejor a la naturaleza de los datos, su empleo en la industria local, no es frecuente. Las dificultades para su implementación pueden tener su causa fundamental en las actividades y cálculos más complejos que ellos requieren .

Sólo algunos programas especializados, entre los que se pueden mencionar el SIMCA [1] y PLS Toolbox de MATLAB[2], permiten aplicar en forma rápida y sencilla esta metodología. Los programas estadísticos de mayor difusión tienen implementada una amplia variedad de procedimientos gráficos útiles para el control univariado de procesos, pero no tienen, en el mismo grado de desarrollo, la que correspondería al control multivariado, como gráficos T^2, y otros.

Es por ello que una mayor divulgación de los métodos de control multivariado y su implementación en programas de computación accesibles a los usuarios, constituirán avances decisivos en la explotación de esta metodología para la mejora de la calidad.

Bibliografía

-Fuchs, C. y Kenett, R : "Multivariate Qulity Control- Theory and Applications". Marcel Dekker Inc., 1998.

-Prins, J. y Mader, D : "Multivariate Control Charts for grouped and individual observations". Quality Engineering, 10(1), 49-57 (1997-98).

[1] SIMCA. UMETRICS AB. Suecia.

[2] PLS_Toolbox Version 1.5. Eigenvector Technologies. Manson, EA