

Gráficos de Control Multivariantes

KEVIN GARCÍA^{1,a}, ALEJANDRO VARGAS^{1,b}

¹DEPARTAMENTO DE ESTADÍSTICA, UNIVERSIDAD DEL VALLE, CALI, COLOMBIA

Resumen

El control de procesos multivariado es una de las secciones de desarrollo de control estadístico de más rápido desarrollo. Hoy en día, en la industria, hay muchas situaciones en las que es necesario el monitoreo o control simultáneo de dos o más características relacionadas con la calidad del proceso. Los problemas de monitoreo de procesos en los que varias variables relacionadas son de interés se conocen colectivamente como Control Estadístico Multivariante (MSPC). En este documento veremos la descripción y aplicación de los gráficos de control multivariados más utilizados en la práctica, por medio de los cuales se quiere conocer si un proceso está bajo control o si, por el contrario, éste se encuentra fuera de control y así, poder centrarse en qué punto el proceso se salió de control, y cuál o cuáles de las variables fueron las causantes de este inconveniente, para corregir el problema de una manera eficiente y continuar satisfactoriamente con el proceso.

1. Introducción

Durante la Segunda Guerra Mundial la técnica estadística de gráficos de control univariados fue la más utilizada a pesar de que los procesos y productos que se analizaban poseían en su gran mayoría dos o más características de calidad (Luis, 2010). Con el tiempo se dio la necesidad de aplicar herramientas estadísticas multivariadas para controlar en forma simultánea dos o más variables (Herrera & Fontalvo, 2006; Miguel, 2010; Marta, 2016). Sin embargo, las técnicas multivariadas de control son técnicas muy complejas de utilizar, por los conceptos matemáticos que se manejan. Esta dificultad es superada posteriormente con el avance de los programas o software especializados en control estadístico de procesos, lo que originó un interés de las técnicas de control multivariado.

2. Antecedentes

Para estudiar un poco cómo funcionan y en qué contextos se han aplicado los gráficos de control multivariantes se estudiaron tres artículos, el primero, **Aplicación del control estadístico multivariado para medir la capacidad del proceso de fabricación de resortes de compresión en acero inoxidable** escrito por **Herrera et al. (2018)** donde se realizó la carta de control multivariada para determinar los parámetros del proceso y luego se calculó el índice de capacidad multivariado para revisar cumplimiento. Los resultados permitieron evidenciar que el proceso en la fase inicial estaba en control y era capaz de cumplir las especificaciones, pero luego en la segunda fase se encontró un proceso con alta variabilidad y poca capacidad de cumplimiento. El segundo artículo, **Desarrollo de una aplicación para gráficos de control de procesos industriales** escrito por **Sánchez (2013)** donde se crea una aplicación para el control estadístico de la calidad basado en un entorno gráfico cuyo motor de cálculo es el lenguaje R. El nombre de la librería es qcr que corresponden a las siglas Quality Control and Reability en la cual se implementan los gráficos de control de calidad más utilizados en este campo. Por último, **Arbeláez (2013)** en su artículo **Gráficos de control no paramétricos basados en R-estadísticos con aplicación al caso multivariante** propone la implementación en R de varios gráficos de control para procesos industriales. Se hace énfasis en los gráficos de control no paramétricos basados en rangos ordinales, tanto univariantes como multivariantes propuestos por Liu y realiza una comparación entre estos y los gráficos de control clásicos propuestos por Shewhart.

^aUniversidad del Valle. E-mail: kevin.chica@correounivalle.edu.co

^bUniversidad del Valle. E-mail: jose.alejandro.vargas@correounivalle.edu.co

3. Descripción metodología de aplicación

Las gráficas de control multivariadas son un tipo de gráfica de control de variables que muestra cómo las variables correlacionadas o dependientes influyen en forma conjunta en un proceso o resultado. Por ejemplo, puede utilizar una gráfica de control multivariada para investigar si la temperatura y la presión se encuentran bajo control en la producción de piezas plásticas moldeadas por inyección.

Si los datos incluyen variables correlacionadas, entonces crear gráficas de control separadas para cada variable podría conducir a una interpretación errónea, debido a que las variables en conjunto afectan el proceso. Si usted utiliza gráficas de control univariadas separadas en una situación multivariada, el error de tipo I y la probabilidad de que un punto caiga correctamente dentro de los límites de control no es igual a sus valores esperados. La distorsión de estos valores aumenta con el número de variables de medición.

Sin embargo, las gráficas de control multivariadas son más difíciles de interpretar que las gráficas de control clásicas de Shewhart. Por ejemplo, la escala de las gráficas de control multivariadas no está relacionada con la escala de ninguna de las variables. Además, las señales de fuera de control en gráficas de control multivariadas no revelan cuál variable (o combinación de variables) causó la señal.

Los gráficos de control multivariantes más utilizados y los cuales se tratarán en este artículo son:

- Gráfico de contorno y gráfico de control χ^2
- Gráfico T^2 o de Hotelling
- Gráfico de control de varianza generalizada
- Gráfico multivariante de promedios ponderados exponencialmente (MEWMA)
- Gráfico multivariante de suma acumulativa (MCUSUM)

3.1. Gráfico de contorno y gráfico de control χ^2

En la distribución normal multivariable, la densidad se describe mediante un elipsoide centrado en el vector medio con ejes en dirección a los vectores propios (e) de la matriz de covarianza, estableciendo μ como el origen y con la longitud

$$\pm c\sqrt{\lambda_j}e_j$$

siendo

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

Si x se distribuye $N_p(\mu, \Sigma)$, entonces $(x - \mu)' \Sigma^{-1} (x - \mu)$ es $\chi_{\alpha, p}^2$

Por lo tanto,

$$(x - \mu)' \Sigma^{-1} (x - \mu) \leq \chi_{\alpha, p}^2$$

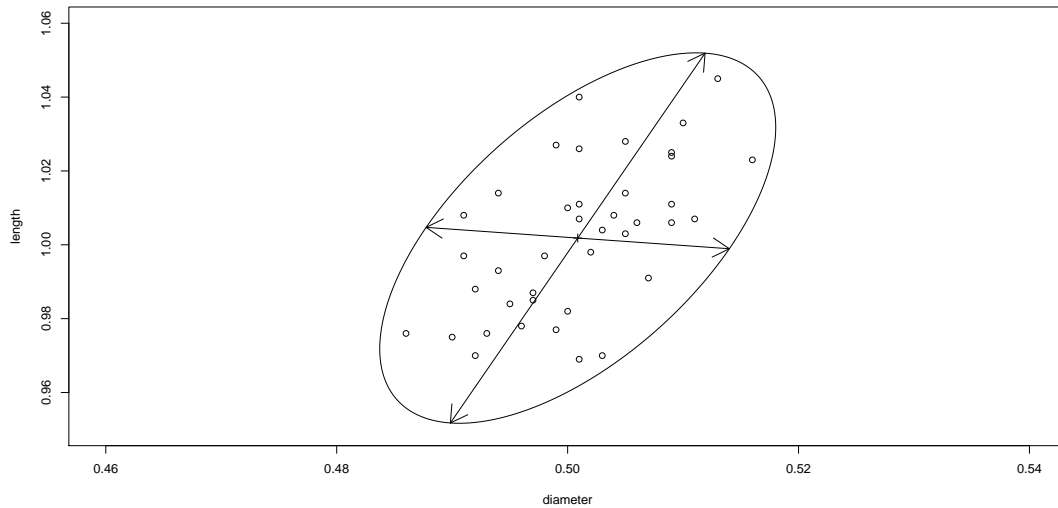


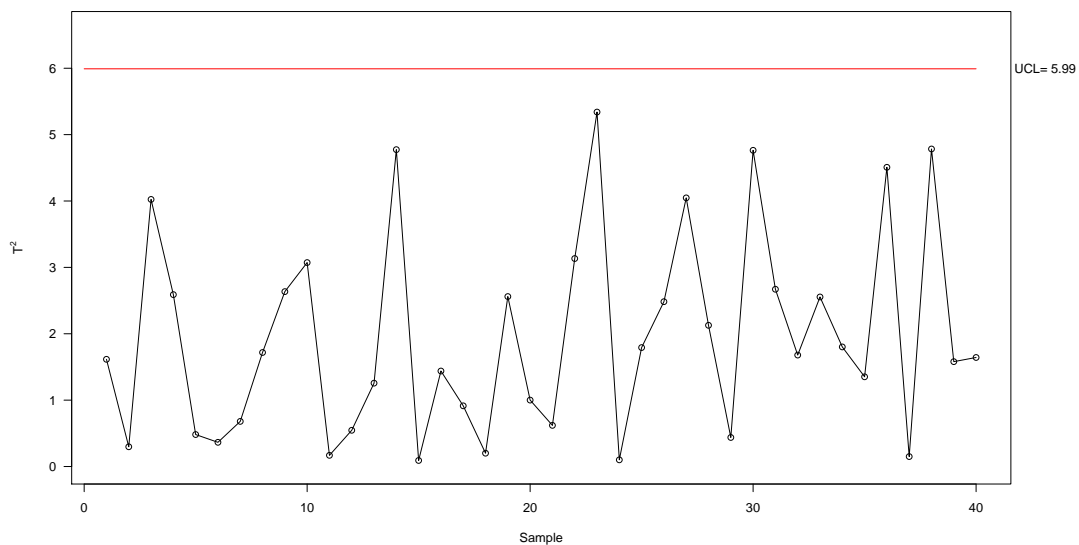
FIGURE 1: Ejemplo gráfico de contorno

Al no obtener puntos fuera de la elipse, no hay evidencia de causas especiales; por lo tanto el proceso está en control. La dificultad para identificar los puntos más allá del elipsoide de confianza es uno de los principales inconvenientes de la herramienta. Otra desventaja es la complejidad para construir el elipsoide cuando $p > 2$, que se puede resolver utilizando el gráfico de control χ^2 que se obtiene al trazar las estadísticas de la prueba:

$$n(x - \mu)'(\Sigma)^{-1}(x - \mu) = \chi^2_{\alpha,p}$$

Donde n es el tamaño de muestra y el límite de control superior:

$$UCL = \chi^2_{\alpha,p}$$

FIGURE 2: Ejemplo gráfico de control χ^2

Mostrando resultados iguales al control elipsoide. Una ventaja de este gráfico es que permite ver la evolución de las muestras a lo largo del tiempo.

3.2. Gráfico T^2

El procedimiento de Hotelling (1947) se ha convertido, sin duda, en el más aplicado en el control de procesos multivariado y es el análogo multivariado del gráfico de control de Shewhart. Por esa razón, también se conoce como gráfico de control Shewhart multivariado. En la práctica, los parámetros μ y Σ son desconocidos y, por lo tanto, deben estimarse a través de los estimadores no sesgados \bar{x} y S . En base a la generalización multivariable del estadístico t a partir de la teoría normal univariada:

$$t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$$

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

por lo que la generalización resulta en

$$T^2 = n(\bar{X} - \bar{\bar{X}})'(S)^{-1}(\bar{X} - \bar{\bar{X}})$$

con \bar{X} y S vector de medias y matriz de covarianza, respectivamente. El estadístico T^2 sigue una distribución F con p y $(mn - m - p + 1)$ grados de libertad. Por lo tanto, para establecer el control en la Fase I, los resultados de UCL en

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

Mientras para monitorear futuras observaciones (Fase II), el límite está dado por

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

Aquí, el número de muestras (m) se refiere a las muestras preliminares tomadas para establecer el estado de control (Fase I). Este gráfico carece de límites de control inferiores (LCL) de manera análoga al gráfico χ^2 .

Este gráfico se emplea en estudios multivariados introductorios y tiene un buen desempeño en la detección de grandes cambios en la media. Según Lowry y Montgomery (1995), la aplicación de este gráfico requiere varias características de calidad entre 2 y 10, y toma más de 20 muestras (a menudo más de 50) de tamaño 2, 3 o 10. Estos valores a veces están limitados por la misma naturaleza del problema.

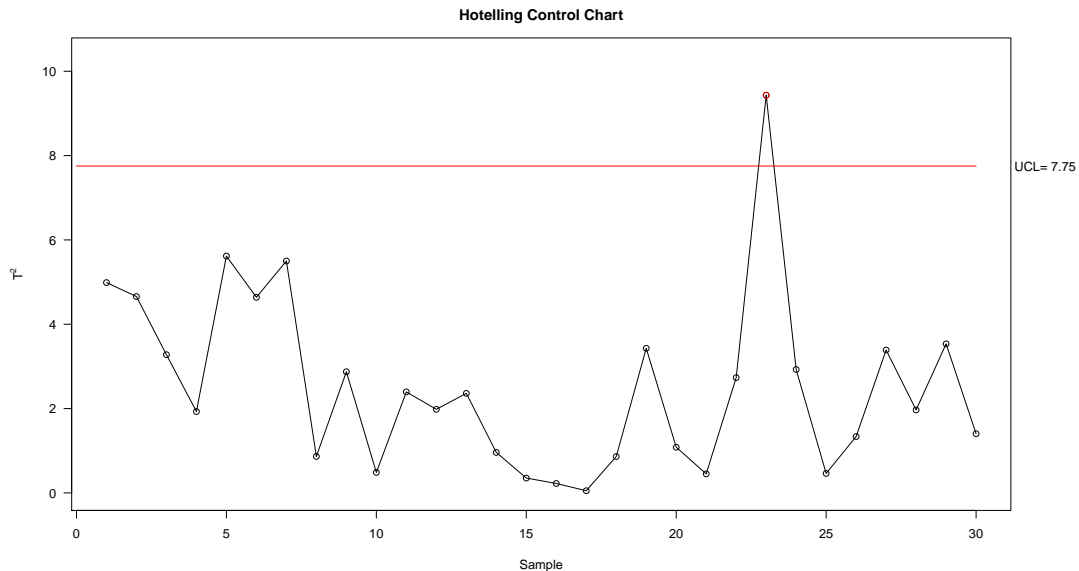


FIGURE 3: Ejemplo gráfico de control T^2 de Hotelling

3.3. Gráfico de control de varianza generalizada

De la misma manera que en el gráfico de control univariado, el monitor de la media del proceso está acoplado con un gráfico de dispersión; El seguimiento de la variabilidad del proceso resulta extremadamente útil en problemas multivariados. Esto se debe a que en el gráfico multivariado de Shewhart se asumió que la dispersión del proceso se mantuvo constante. Esta hipótesis debe comprobarse en la práctica.

Hasta la fecha, se han propuesto varios métodos para el monitoreo simultáneo de la variabilidad, pero claramente el gráfico de varianza generalizada es el más aceptado. El término varianza generalizada se conoce como el determinante de la matriz de covarianza.

Este tipo de gráfico resulta al trazar el determinante de la matriz de covarianza junto con los límites de control superior e inferior naturales. Cuando se conoce la matriz de covarianza S , los parámetros del gráfico dan como resultado

$$UCL = |\Sigma|(b_1 + 3b_2^{1/2}) \quad ; \quad CL = b_1|\Sigma| \quad ; \quad LCL = \max\{|\Sigma|(b_1 - 3b_2^{1/2}), 0\}$$

Donde

$$b_1 = \frac{1}{(n-1)^p} \prod_{j=1}^p (n-j)$$

$$b_2 = \frac{1}{(n-1)^{2p}} \prod_{j=1}^p (n-j) \left[\prod_{i=1}^p (n-i+2) - \prod_{i=1}^p (n-1) \right]$$

Tenga en cuenta que n debe ser mayor que el número de características de calidad (p).

Con frecuencia, Σ se desconoce y se estima a través de S según la relación:

$$|S| = b_1|\Sigma|$$

Por lo tanto, los parámetros resultan en

$$UCL = \frac{S}{b_1}(b_1 + 3b_2^{1/2}) \quad ; \quad CL = |S| \quad ; \quad LCL = \max\left\{\frac{|S|}{b_1}(b_1 - 3b_2^{1/2}), 0\right\}$$

Teniendo en cuenta que S es una matriz positiva definida, el LCL carece de sentido para los valores negativos.

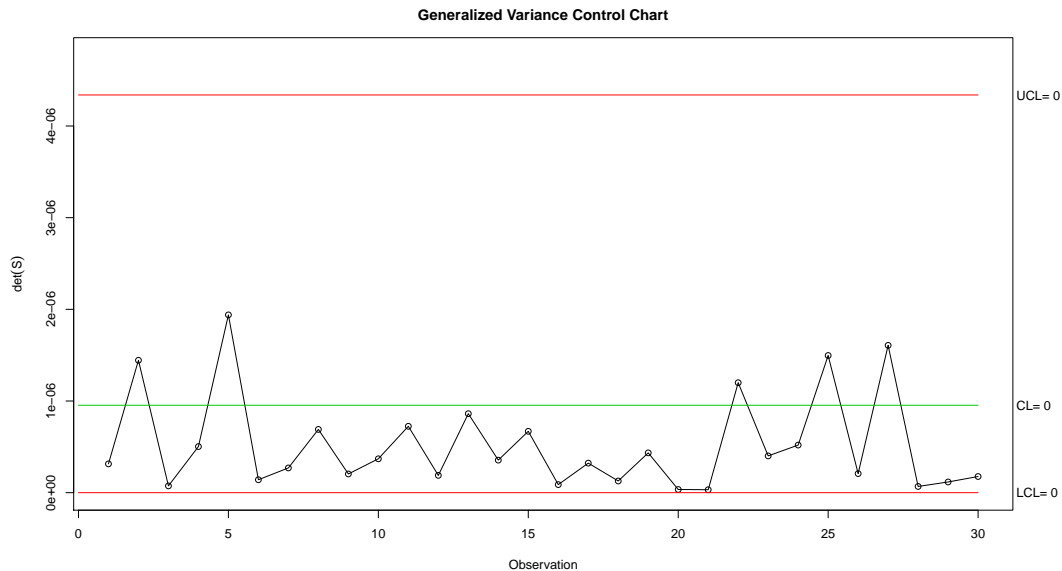


FIGURE 4: Ejemplo gráfico de control de varianza generalizada

3.4. Gráfico multivariante de promedios ponderados exponencialmente MEWMA

MEWMA es la extensión multivariable natural del gráfico EWMA propuesto por Roberts (1959). Fue introducido por Lowry et al. (1992) y es más sensible en la detección de cambios no aleatorios en el proceso y se basa en el principio del promedio ponderado de los vectores observados anteriormente. El gráfico MEWMA tiene las estadísticas:

$$T^2 = Z_i' \Sigma_{Z_i}^{-1} Z_i > h$$

Donde

$$Z_i = \lambda X_i + (1 - \lambda) X_{i-1}$$

Siendo $Z_0 = 0$, λ es una matriz diagonal $p \times p$ de constantes de suavizado con $0 < \lambda_i \leq 1$, aunque en la práctica no hay razón para emplear valores diferentes de λ en el mismo problema. En la práctica, el valor más utilizado de λ es 0.1. En un caso particular, cuando se obtienen subgrupos racionales, es decir, $n > 1$, simplemente reemplace X_i por \bar{X}_i .

Lowry et al. (1992) proporcionan dos alternativas para calcular el tamaño, la matriz de covarianza exacta:

$$\Sigma_{Z_i} = \frac{\lambda [1 - (1 - \lambda)^{2i}]}{2 - \lambda} (\Sigma)$$

Y, la llamada matriz de covarianza asintótica:

$$\Sigma_{Z_i} = \frac{\lambda}{2 - \lambda} (\Sigma)$$

El primero teniendo un mejor desempeño. Además, señalan que el rendimiento de ARL del gráfico depende solo del parámetro de no centralidad θ :

$$\theta = [(\mu_1 - \mu_0)' \Sigma (\mu_1 - \mu_0)]^{1/2}$$

Donde μ_1 es el vector de medias para la fase II. Note que cuando $\lambda = 1$ el gráfico MEWMA es transformado en el gráfico T^2 .

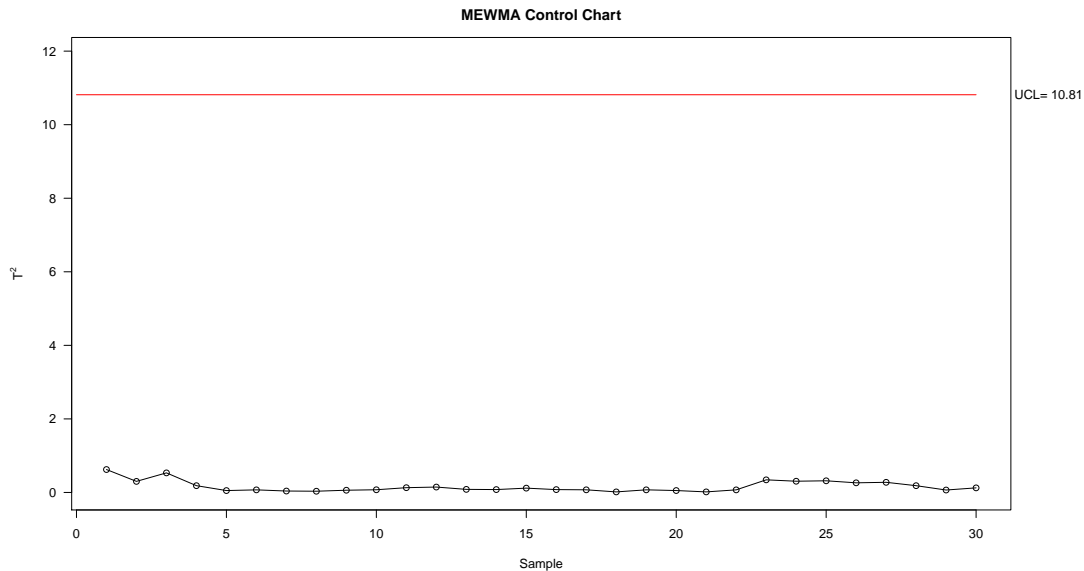


FIGURE 5: Ejemplo gráfico de control multivariante de promedios ponderados exponencialmente MEWMA

3.5. Gráfico multivariante de suma acumulativa MCUSUM

El gráfico de control MCUSUM aparece como la extensión multivariada del gráfico de control CUSUM propuesto originalmente por Page (1961). Se enfoca en mejorar la sensibilidad con respecto al gráfico T^2 previamente introducido al detectar pequeños cambios en el proceso y se basa en el principio de acumular información de las observaciones anteriores. Además del gráfico MEWMA, MCUSUM es un gráfico de la Fase II. Hay cuatro alternativas principales aceptadas para construir un gráfico MCUSUM, por limitaciones en la extensión nos enfocaremos en las dos propuestas que están implementadas en R, MCUSUM de acuerdo a Crosier (1988) y MCUSUM de Pignatiello y Runger (1990).

3.5.1. MCUSUM Crosier (1988)

Crosier (1988) presentó dos procedimientos multivariados. Se presenta la versión que tiene un mejor desempeño. El estadístico es:

$$T_i^2 = \left[S_i' \left(\frac{\Sigma}{n} \right)^{-1} S_i \right]^{1/2} > h$$

Donde

$$S_i = \begin{cases} 0 & \text{si } C_i \leq k \\ (S_{i-1} + \bar{X}_i - \mu_0) \left(1 - \frac{k}{C_i} \right) & \text{si } C_i > k \end{cases}$$

Donde $S_0 = 0$, $k > 0$, y

$$C_i = \left[(S_{i-1} + \bar{X}_i - \mu_0)' \left(\frac{\Sigma}{n} \right)^{-1} (S_{i-1} + \bar{X}_i - \mu_0) \right]^{1/2}$$

Asimismo, el límite es

$$UCL = h$$

3.5.2. MCUSUM Pignatiello y Runger (1990)

Pignatiello y Runger también propusieron dos gráficos, el siguiente es el que tuvo mejor rendimiento:

$$T_i^2 = \max \begin{cases} 0 \\ \left[S_i' \left(\frac{\Sigma}{n} \right)^{-1} S_i \right]^{1/2 - kn_i} \end{cases}$$

Donde

$$S_i = \sum_{j=1-n_i+1}^i (\bar{X}_j - \mu_0)$$

Y

$$n_i = \max \begin{cases} n_{i-1} + 1 & \text{si } T_{i-1}^2 > 0 \\ 1 & \text{En otro caso} \end{cases}$$

$$UCL = h$$

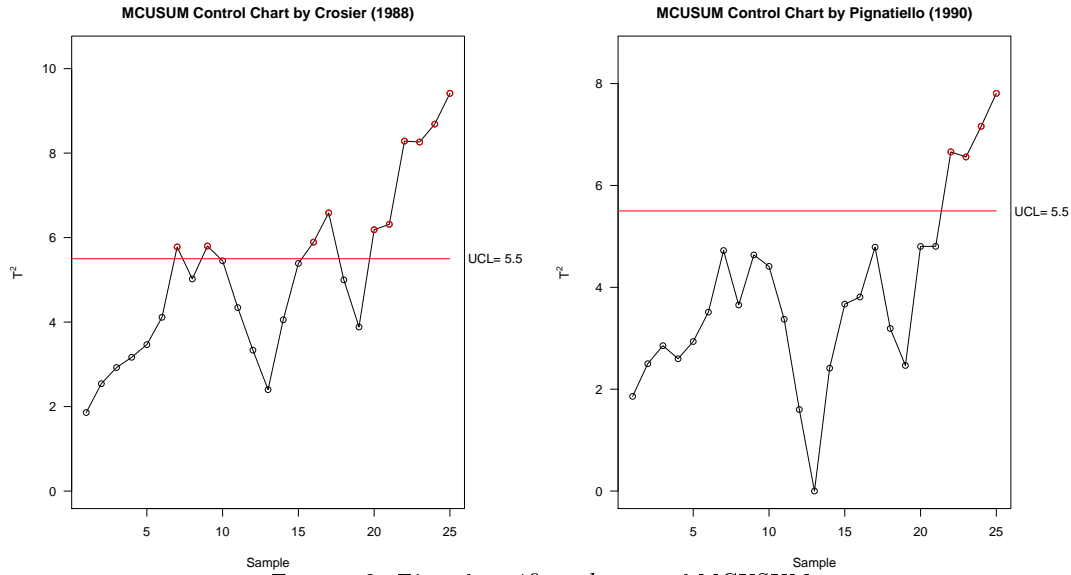


FIGURE 6: Ejemplo gráficos de control MCUSUM

4. Caso ilustrativo: Control de pitcheo

En este caso ilustrativo, la aplicación de los gráficos de control multivariados se introducen en el béisbol utilizando el software (R Core Team 2019), específicamente sobre el desempeño del lanzador. Si bien los lanzadores mueven la pelota estratégicamente en diferentes posiciones de la zona de ataque intentando que el bateador no haga contacto con ella, a menudo el desempeño del lanzador se mide por la habilidad para poner la bola en la zona de golpe a alta velocidad.

En este caso, utilizamos los datos recopilados por los registros de lanzadores en la base de datos (MLB) (<http://gd2.mlb.com/components/game/mlb/>) para el lanzador C.C. Sabathia de los New York Yankees. Se seleccionaron dos conjuntos de datos de los juegos contra Tampa Bay: el primero el 10 de julio de 2011 y el 12 de agosto de 2011 el segundo. Ambos se almacenan en el paquete (Santos-Fernández 2013) como `sabathia1` y `sabathia2` respectivamente. Los registros del lanzador brindan mucha información sobre cada lanzamiento, pero en nuestro estudio trabajamos con la velocidad de inicio (dada en mph) del lanzamiento y la ubicación (en pies) cuando cruza la casa. Este último punto se mide con respecto a un sistema de coordenadas en el que el origen se encuentra en el punto de la placa base. El eje z es el orientado verticalmente mientras que el eje x orientado horizontalmente a la derecha del receptor. Solo se consideran los lanzamientos de bola rápida y cada muestra es un bateador promediando todas las variables de lanzamiento. Observa que un jugador bate varias veces en la jugada.

El vector de medias, la matriz de varianzas y covarianzas y la matriz de correlaciones son:

$$\bar{x} = \begin{bmatrix} 0.1069565 \\ 2.9430435 \\ 94.4108696 \end{bmatrix}; S = \begin{bmatrix} 0.22087668 & 0.09182787 & 0.05130731 \\ 0.09182787 & 0.27479486 & -0.25359822 \\ 0.05130731 & -0.25359822 & 1.50752648 \end{bmatrix}; r = \begin{bmatrix} 1.00000000 & 0.3727308 & 0.08891435 \\ 0.37273078 & 1.0000000 & -0.39401198 \\ 0.08891435 & -0.3940120 & 1.00000000 \end{bmatrix}$$

Se puede realizar un análisis útil inicial mediante la construcción de un diagrama de dispersión tridimensional con un elipsoide de confianza. Al moverse a través de las coordenadas, se puede ver

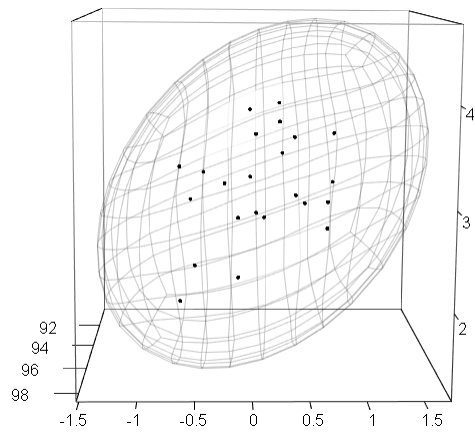
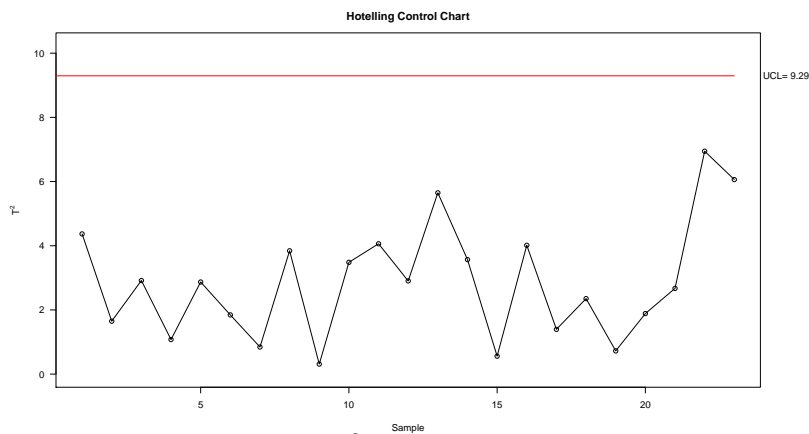


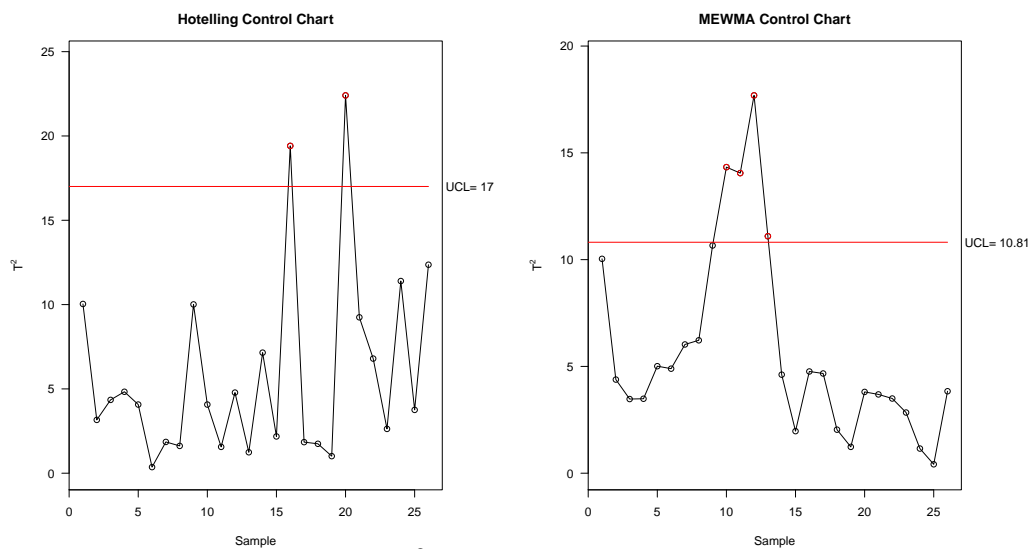
FIGURE 7: Diagrama de dispersión tridimensional con confianza elipsoide

que todas las observaciones caen dentro de estos límites. No se detectan valores atípicos. Al realizar un gráfico de Hotelling, obtenemos

FIGURE 8: Gráfico de control T^2 de Hotelling para los datos de sabathia 1

Dado que no hay puntos fuera del UCL, no hay evidencia para rechazar el estado de control en el proceso. La puntuación final lo demuestra.

Luego, Se utiliza este primer juego para analizar el segundo juego como Fase II u observaciones futuras, utilizando las estimaciones de Fase I del vector medio y la matriz de covarianza.

FIGURE 9: Gráficos de control T^2 de Hotelling y MEWMA para los datos sabathia2

Para detectar cuál o cuáles variables originaron que el proceso se saliera de control, y poder corregir el problema, se utiliza un proceso denominado descomposición del estadístico T^2 , la cual obtiene el estadístico T^2 y el UCL con cada una de las variables individuales, luego con pares de variables para las posibles combinaciones y finalmente para las 3 variables. La variable o variables para las cuales el T^2 calculado supera el valor del UCL calculado son las que originaron que el proceso se saliera de control. La descomposición del estadístico T^2 para los dos puntos (16 y 20) o muestras en los cuales el proceso se sale de control es:

\$*Decomposition of*	\$*Decomposition of*
[1] 16	[1] 20
t2 decomp ucl p-value 1 2 3	t2 decomp ucl p-value 1 2 3
[1.] 12.5255 8.0686 0.0016 1 0 0	[1.] 0.4091 8.0686 0.5282 1 0 0
[2.] 10.7037 8.0686 0.0031 2 0 0	[2.] 9.6004 8.0686 0.0048 2 0 0
[3.] 4.2001 8.0686 0.0511 3 0 0	[3.] 0.8664 8.0686 0.3609 3 0 0
[4.] 16.8950 12.1448 0.0000 1 2 0	[4.] 13.4175 12.1448 0.0001 1 2 0
[5.] 18.1565 12.1448 0.0000 1 3 0	[5.] 1.3922 12.1448 0.2671 1 3 0
[6.] 11.3942 12.1448 0.0003 2 3 0	[6.] 15.0562 12.1448 0.0001 2 3 0
[7.] 19.4116 16.1352 0.0000 1 2 3	[7.] 22.4067 16.1352 0.0000 1 2 3

FIGURE 10: Descomposición estadístico T^2 para detectar variables fuera de control

El análisis muestra los puntos 16 y 20 más allá de la UCL , es decir: el lanzador parece estar fuera de control. Teniendo en cuenta que las variables 1,2 y 3 son, respectivamente, eje horizontal, eje vertical y velocidad del lanzamiento, la descomposición de las estadísticas T^2 indica como en la muestra 16 ambas ubicaciones en los ejes horizontal y vertical (x) estaban fuera de control. En contraste, en el bateador número 20 solo la ubicación en la vertical causa la alarma.

Con el fin de mejorar la detección rápida de pequeños cambios en el proceso, se calcularon los gráficos MEWMA y MCUSUM. Se puede observar, que el gráfico MEWMA de la figura 9 detecta los cambios en la media en el 10º bateador.

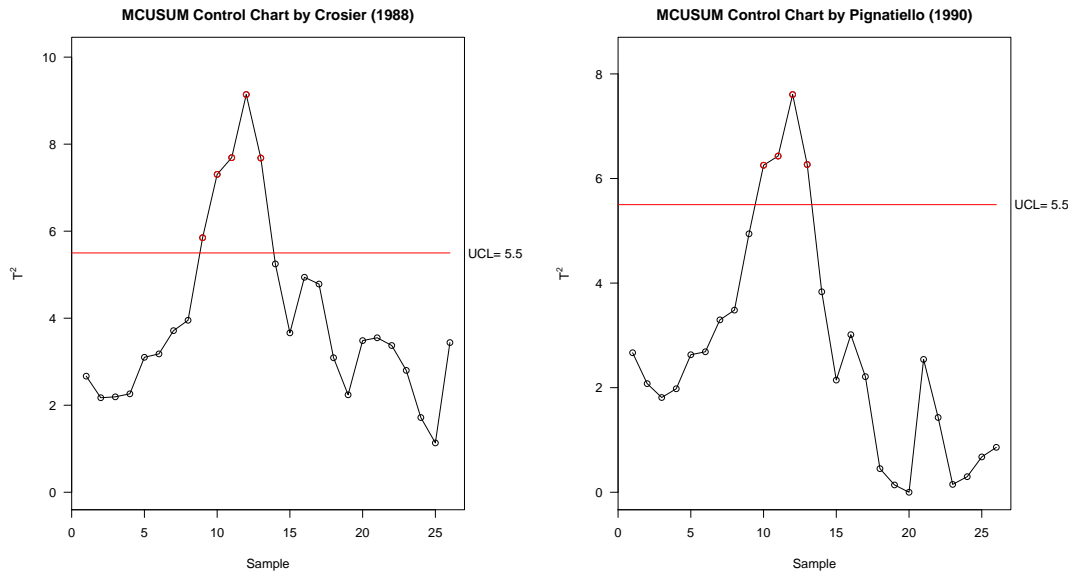


FIGURE 11: Gráfico de control CUSUM por Crosier 1988 y Pignatiello 1990 para los datos de sabathia2

El MCUSUM según (Crosier 1988) y (Pignatiello y Runger 1990) en los 9º y 10º bateadores respectivamente.

5. Conclusiones

Como conclusión general se tiene que los gráficos de control multivariantes son una herramienta muy útil y muy potente en el control estadístico de la calidad, permitiéndonos detectar cuando un proceso se sale de control con una alta precisión. Dentro de los gráficos mencionados, se vio que el gráfico de control T^2 de Hotelling tiene un buen desempeño en la detección de grandes cambios en la media, sin embargo, se queda corto ante cambios pequeños que presente el proceso; por ello se utilizan

también los gráficos MEWMA y MCUSUM los cuales mejoran la detección rápida de pequeños cambios en el proceso, en el ejemplo aplicado se vio que estos últimos detectaban más rápido los cambios en el desempeño de los lanzamientos del jugador.

La importancia de estos gráficos de control mutivariados radica en tres principales aspectos; representan la región de control real de las variables relacionadas (elíptico para casos bivariados), permiten mantener la tasa especificada de error tipo I y permiten monitorear todas las variables del proceso correlacionadas en una sola gráfica, frecuentemente con un solo límite de control.

References

- Allen, T. T. (2006), *Introduction to Engineering Statistics and Six Sigma*.
- Arbeláez, R. D. L. (2013), ‘Gráficos de control no paramétricos basados en r estadísticos con aplicación al caso multivariante’.
- Fernández, E. S. (2012), *Multivariate Statistical Quality Control Using R*.
- Herrera, R., Hernández, K., Figueroa, E. & Ávila, J. D. (2018), ‘Aplicación del control estadístico multivariado para medir la acapacidad del proceso de fabricación de resortes de compresión en acero inoxidable’, *Prospectiva* **16**(2), 49–58.
- R Core Team (2019), *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria.
*<https://www.R-project.org/>
- Sánchez, M. A. F. (2013), Desarrollo de una aplicación para gráficos de control de procesos industriales, Master’s thesis, Universidad de la Coruña.
- Santos-Fernández, E. (2013), *Multivariate Statistical Quality Control Using R*, Vol. 14, Springer.
*<http://www.springer.com/statistics/computational+statistics/book/978-1-4614-5452-6>