Entre Ciencia e Ingeniería, ISSN 1909-8367
Año 6. No. 11 – Primer semestre de 2012, páginas 164-182

Comparando la Eficiencia de los Gráficos T^2 de Hotelling y de Clasificación por Rangos, Utilizando Diversos Estimadores de Localización y Dispersión[1]

Comparing the Efficiency of T^2 Graphs for Hotelling and Classification by Ranking, Using Different Estimators of Location and Dispersion

Joaquín González Borja
Profesional en Matemáticas con énfasis en Estadística
Maestría en ciencias Estadística
Docente Departamento de Matemáticas y Estadística Universidad del Tolima
Grupo de investigación GIESTUT
jgonzalezb@ut.edu.co

Haider Montes Masmela
Estudiante Matemáticas con énfasis en Estadística
Universidad del Tolima
haimon17@hotmail.com

Oscar Andrés Lugo Capera
Estudiante Matemáticas con énfasis en Estadística
Universidad del Tolima
lugynhnio@gmail.com

Recibido Septiembre 30 de 2011 – Aceptado Mayo 30 de 2012

RESUMEN

Para muchos productos o servicios, su calidad es definida simultáneamente por un conjunto de variables correlacionadas. Para monitorear un proceso multivariado, el gráfico de control clásico (Hotelling (1947)) es

[1] Documento derivado del proyecto terminado "Evaluación de la efectividad de la descomposición MYT y del gráfico de contribuciones en la identificación de variables causantes de señales fuera de control en cartas de control T^2 y de rangos robustas", Código 200110 del centro de investigaciones de la Universidad del Tolima.

frecuentemente usado. Este gráfico es construido bajo el supuesto de normalidad de las observaciones y con estimadores de localización y dispersión usuales. Es bien conocido, que este gráfico es muy sensible a la presencia de outliers en el *conjunto de datos históricos provocando el efecto de enmascaramiento, por lo cual han surgido varias propuestas de construcción de gráficos con estimadores alternativos que los ha convertido en gráficos más potentes y robustos ante la presencia de datos atípicos o en la detección más rápida del cambio en el vector de medias. La normalidad de las observaciones no siempre se cumple en la práctica, luego los gráficos de control no paramétricos son los recomendados en este caso, uno de ellos es el gráfico de clasificación por rangos (Liu (1995)). Trabajos como el de Zertuchi y Cantú (2008), Velásquez y Moreno (2009), comparan la eficiencia de los dos gráficos mencionados usando estimadores usuales, bajo normalidad y carente de ella.*

En este trabajo, se presenta un estudio comparativo de la eficiencia de los dos gráficos usando diversos estimadores del vector de medias y de la matriz de covarianzas, en presencia de datos atípicos en la fase de construcción. La eficiencia de los gráficos de control, es determinada por la pronta detección en el cambio del vector de medias. La eficiencia de los gráficos en este estudio, se mide a través del cálculo de la longitud promedio de corrida (ARL) bajo control, mediante simulación estadística ante diversos ambientes planteados. Se presentan los resultados obtenidos con sus respectivas interpretaciones y conclusiones.

Palabras Clave: Gráficos de Control Multivariado, Estadística T^2, Clasificación por Rangos, Estimadores de Localización y Dispersión, Longitud Promedio de Corrida.

ABSTRACT

Quality is defined simultaneously by a set of correlated variables for many products or services. To monitor a multivariate process, the classic control chart T^2 (Hotelling (1947)) is often used. This chart is constructed under the assumption of normality of the observations and estimates of usual location and dispersion. It is well known that this chart is very sensitive to the presence of outliers in the historical data set, causing the masking effect and therefore several proposals to construct charts T^2 with alternative estimators that have become more powerful and robust charts in the presence of outliers or faster detection of change in the mean vector have been proposed.

The normality of the observations is not always true in practice, then the nonparametric control charts are recommended in this case, one of them is the chart of a ranking method (Liu (1995)). Zertuchi and Cantu (2008), Velasquez and Moreno (2009), comparing the efficiency between the two charts above using conventional estimators under normal and devoid of it. This work, presents a comparative study about the efficiency of the two charts using various estimates of the mean vector and covariance matrix, in the presence of outliers in the construction phase. The efficiency of control charts is determined by the change in the early detection of mean vector. The efficiency of charts in this study is measured calculating the average run length (ARL) under control by statistical simulation through various environments encountered. We present the results obtained with their respective interpretations and conclusions.

Key words: Multivariate Control Charts, Statistics T^2, Rankings Index, Estimates of Location and Dispersion, Average Run Length.

1. INTRODUCCIÓN

El rápido crecimiento de la tecnología de adquisición de datos y el uso de computadores en línea para el monitoreo de un proceso ha incrementado el interés por controlar, en forma simultánea, diversas características de calidad de un mismo producto o servicio y esto ha llevado a estudiar las técnicas donde se consideren procedimientos de control estadístico del proceso en forma multivariada.

Diversas herramientas para el monitoreo de las características de calidad de un producto o servicio en el contexto multivariado aparecen reportadas en la literatura, desde la técnica establecida por Hotelling (1947), denominado el gráfico de control T^2, pasando por gráficos CUSUM y EWMA multivariadas, el uso de componentes principales, mínimos cuadrados parciales, redes neuronales, modelos no lineales, métodos no paramétricos entre otros; con el propósito de monitorear la media y/o dispersión del proceso.

El gráfico T^2 es la herramienta más utilizada para el monitoreo del vector de medias de varias características de calidad correlacionadas, dada su facilidad de cálculo e interpretación. Pero esta, es muy poco sensible a cambios pequeños en el vector de medias y pierde potencia y robustez en la fase de monitoreo (fase II) ante la presencia de observaciones atípicas (outliers) en el conjunto de datos históricos los cuales se usan para su

construcción (fase I), estimando el vector de medias μ y la matriz de covarianzas $\sum$.

Esta problemática ha llevado al uso de estimadores alternativos, entre los que se cuentan las propuestas de Sullivan y Woodall (1996), Vargas (2003), Willems y Cols. (2002) y Chenouri y Cols. (2008) quienes usaron estimadores de la covarianza basados en diferencias sucesivas entre las observaciones (SW), estimadores de elipsoide de volumen mínimo (MVE), estimadores de la covarianza de determinante mínimo (MCD) y estimadores de covarianza de determinante mínimo reponderado (RMCD), respectivamente. Sus resultados sugieren un mejoramiento en la eficiencia del gráfico cuando este ha sido construido bajo observaciones individuales.

Otro de los inconvenientes que se puede presentar en la pérdida del buen funcionamiento del gráfic T2 es la carencia de normalidad multivariada en las observaciones, lo cual suele ocurrir con mucha frecuencia en la práctica. Para enfrentar la situación, se han desarrollado gráficos de control no paramétricos, uno de ellos es el propuesta por Liu (1995) llamado gráfico de comparación de rangos, basado en el concepto de profundidad de datos.

Velásquez y Moreno (2009), presenta en una comunicación la comparación a través de simulación estadística del desempeño de los gráficos T^2 y la de rangos, llegando a la conclusión que el gráfico de rangos tiene un mejor comportamiento en detectar señal tanto para datos normales como no normales. El estudio lo realizaron considerando estimadores usuales para μ y $\sum$.

El propósito de esta investigación es retomar estas temáticas, iniciando con la comparación de las dos metodologías una paramétrica y otra no paramétrica, con el uso de estimadores alternativos en la construcción de los gráficos T^2 y de rangos, asumiendo normalidad en las observaciones. Dicho estudio se realiza vía simulación estadística ante diversos escenarios planeados mediante el cálculo de la longitud promedio de corrida (ARL) de los gráficos; para así, determinar su eficiencia, entendiéndose como eficiencia el gráfico bajo control estadístico, que detecte en una forma más rápida el cambio en el vector de medias ante la presencia de correlación entre las variables de interés y en presencia de datos contaminados.

2. CONSIDERACIONES TEÓRICAS

2.1. Gráficos de Control Multivariado

Hay varias situaciones en las cuales es necesario controlar simultáneamente dos o más características de un producto o proceso. Este tipo de problemas se conoce como control estadístico de procesos multivariado.

De esta forma, se denotará por p el número de características a medir. Interesa conocer cómo usar las medidas simultáneamente, para probar la hipótesis de que el proceso se encuentra bajo control estadísticamente hablando. En este análisis, se deben tener en cuenta las correlaciones entre las características de calidad. Denotemos por X un vector p dimensional que contiene las observaciones de las características de calidad. Se asume que X sigue una distribución $N_p(\mu,\sum)$. Si el valor objetivo de X es μ_0 se necesita establecer un esquema que permita conocer si la media del proceso está estable o no. Es decir, si el proceso está en control o no. Se entiende que el proceso está en control o bajo control cuando la variabilidad respecto al valor objetivo se da debido a causas aleatorias.

Alt y Smith (1988) ha establecido dos fases para el proceso de construcción de un gráfico de control multivariada para la media. La fase I se subdivide a su vez en dos etapas. En la etapa 1 o etapa retrospectiva se prueba si el proceso estaba bajo control cuando se recogieron las primeras observaciones. En la etapa 2 o etapa prospectiva se prueba si el proceso permanece bajo control cuando futuras observaciones sean seleccionadas. En la fase II se usa el gráfico de control para detectar desvíos del proceso respecto de su valor estándar.

El gráfico de control se diseña con base en el hecho de haber llevado a cabo observaciones individuales en cada tiempo o si se tomaron grupos de observaciones en cada punto muestral.

Una aproximación en la construcción de los gráficos de control multivariados se basa en la formación de un estadístico de control único que surge de los datos multivariados en cada una de las muestras. Este estadístico de control puede ser una forma cuadrática que involucra en su construcción, estadísticas de cada variable y se gráfica sobre tipos de cartas Shewhart (Hotelling (1947)). El gráfico resultante tiene la misma desventaja de las cartas Shewhart univariadas, son ineficientes ante cambios pequeños en los parámetros.

Una mejor aproximación es el cálculo de las estadísticas EWMA (Lowry y Cols. 1992)) y varias propuestas para cada una de las variables por la CUSUM (Crosier (1988), Pignatiello y Runger (1990)); se usa una forma cuadrática que combina las estadísticas univariadas en forma separada en una estadística de control única a ser plasmada sobre un gráfico de control usual. Los gráficos mencionados presentan la propiedad de ser invariantes direccionales, es decir dependen de la distancia no de la dirección en el cambio de la media, luego la longitud promedio de corrida (ARL) en los gráficos de control multivariados depende del vector de medias y de la matriz de covarianzas solamente a través del parámetro de no centralidad $\lambda^2(\mu) = (\mu - \mu_0)'\Sigma^{-1}(\mu - \mu_0)$ (Lowry y Montgomery (1995)).

Diversos gráficos de control multivariados con metodología no paramétrica han sido desarrolladas, entre ellos se cuenta el gráfico de comparación de rangos propuesta por Liu (1995), basado en el concepto de profundidad de datos. En este último caso, no es necesario el supuesto de normalidad de las observaciones.

A continuación se describe la construcción y propiedades de los gráficos de tipo multivariado, que se estudiarán en este trabajo: el Gráfico T^2 de Hotelling y la Gráfica de Comparación de Rangos.

2.2. Gráfico T^2 de Hotelling para Observaciones Individuales

En las etapas iniciales del proceso, el vector de medias y la matriz de covarianzas de la población en estudio usualmente se desconocen y deben ser estimados a partir de un conjunto de observaciones tomadas cuando el proceso se encuentra bajo control.

Denotemos por $X_i = (X_{i1}, X_{i2}, \ldots, X_{ip})$ la i-ésima observación, i = 1,2,…, m.

Sean $\overline{X}$ y S la media y la matriz de covarianzas muestrales (estimadores usuales) calculadas a partir de estas m observaciones. Esto es,

$$\overline{X} = \frac{1}{m}\sum_{i=1}^{m} X_i$$

$$S = \frac{1}{m-1}\sum_{i=1}^{m}(X_i - \overline{X})'(X_i - \overline{X}).$$

La estadística T^2 para la i-ésima observación está definida por

$$T_i^2 = \left(X_i - \overline{X}\right)' S^{-1}\left(X_i - \overline{X}\right), \qquad i = 1,2,\ldots,m.$$

En la etapa 1, la distribución de esta estadística es proporcional a una distribución beta:

$$T_i^2 \sim \frac{(m-1)^2}{m} B\left(\frac{P}{2}, \frac{m-P-1}{2}\right)$$

(Wierda (1994), Tracy y Cols. (1992)). El gráfico T^2 es entonces un gráfico de la Estadística T^2 versus el número de observaciones (carta propuesta por Hotelling (1947)). El límite de control superior de este gráfico está dado por

$$LCS = \frac{(m-1)^2}{m} B\left(\frac{P}{2}, \frac{m-P-1}{2}, \alpha\right)$$

donde $B(\delta_1, \delta_2, \alpha)$ es el percentil $(1-\alpha)$ de la distribución beta con parámetros δ_1, δ_2. La desigualdad $T^2 \leq LCS$ describe el interior y la frontera de una elipsoide de dimensión p con centro en X. Si las T_i^2 $(i=1,2,\ldots m.)$ caen dentro de esta elipsoide, entonces se asume que el proceso está estadísticamente bajo control. Si el gráfico T^2 indica una señal, se detiene el proceso de producción y se lleva a cabo una investigación con el objeto de encontrar causas especiales que hubieran producido la señal. Las observaciones multivariadas que correspondan a causas especiales se eliminan. Entonces se calculan nuevos límites de control para un examen retrospectivo basado en las observaciones restantes y el procedimiento se repite de nuevo.

En la etapa 2, para una observación futura X_i la estadística T^2 de Hotelling

$$LCS = \frac{p(m+1)(m-1)}{m(m-p)} F(p, m-p, \alpha)$$

sigue una distribución proporcional a una F. El gráfico a usar en esta etapa es una T^2 con límite de control superior,
donde $F(p,m-p,\alpha)$. es el percentil $(1-\alpha)$ de la distribución F con p y $(m-p)$grados de libertad y m es el número de observaciones con las que finalmente se estimaron los parámetros en la etapa 1.

Una vez se ha estabilizado el proceso en la etapa 2, se asume que los estimadores finales $\overline{X}$ y S son los verdaderos valores de los parámetros bajo control. Así que para la fase II se construye un gráfico X^2 , pues

$$\chi^2 = (X_i - \mu_0)' \Sigma^{-1} (X_i - \mu_0)$$

Sigue una distribución X^2 con p grados de libertad. Entonces

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

y la carta indica una señal si para una observación, $\backslash X_i, \chi_i^2 > \chi^2_{p,\alpha}$.

2.3. Gráfica de Comparación de Rangos (r Chart)

Un enfoque de control no paramétrico para los procesos multivariados fue propuesto inicialmente por Liu (1995), dice que es un gráfico de control, denominado gráfico de Comparación de Rangos, basado en el concepto de profundidad de datos. Esta gráfica detecta en forma simultánea, el cambio en localización y el incremento en escala del proceso. La construcción de esta gráfica es completamente no paramétrica; en particular no requiere la condición de normalidad para la distribución de las características de calidad, tal como ocurre para el buen funcionamiento del gráfico de Hotelling.

Para trabajar con esta metodología bastará con hacer uso de la distribución calculada a partir de los datos históricos, ésta se denominará distribución de referencia, y estará denotada por H, la cual describe una distribución $p(p \geq 1)$ dimensional, donde p son las características de calidad del producto o servicio a monitorear. En este caso el proceso se considerará bajo control si las nuevas observaciones se ajustan a la distribución de referencia. $Y_1, Y_2 \ldots, Y_m$ son m observaciones aleatorias de H.

Sea $X_1, X_2 \ldots, X_l$ una muestra del proceso, se referirá a la distribución de $X_i's$ como distribución de monitoreo, y se denotará por G . Basándose en la distribución de las observaciones $X_i's$ se podrá determinar si el proceso se encuentra o no fuera de control.

Para determinar si las $X_i's$ se ajustan o no a la distribución de referencia H es necesaria comparar las dos distribuciones G y H, se debe asumir que las dos distribuciones son absolutamente continuas. Los estadísticos que se usan para analizar las diferencias entre H y G se basan en la teoría de la profundidad de datos, ésta parte del hecho de que cualquier densidad de probabilidades distingue entre puntos "centrales" y "periféricos", por tanto, cualquier valor de una función de este tipo medirá que tan " profunda" o "central" es y con respecto a H . Una función de profundidad asigna a cada $y \in \mathrm{R}^k$, un rango no negativo, el cual se puede interpretar como

su localización en la nube de puntos, así las profundidades más grandes corresponderán al centro de la distribución y las más pequeñas a las regiones externas.

Las funciones de profundidad de datos deben cumplir con las propiedades de ser invariante afín, maximilidad al centro y desvanecimiento al infinito (Zuo y Serling (2000)). De estas funciones se tiene entre otras, la simplicial, Tukey, Majority y la distancia de Mahalanobis, esta última una de las más usadas. La función de Mahalanobis se define como:

$$MD_H(y) = \frac{1}{[1 + (y - \mu_H)' \ \Sigma_H^{-1}(y - \mu_H)\ ]} \qquad (1)$$

donde μ_H y $\sum_H^{-1}$ son el vector de medias y la matriz de covarianzas de la distribución de referencia H, respectivamente; y es el vector de datos observados en las características de calidad analizadas. Si los parámetros de la distribución son desconocidos, se tiene la siguiente ecuación:

$$MD_H(y) = \frac{1}{[1 + (y - \overline{y})' S^{-1}(y - \overline{y})\ ]} \qquad (2)$$

Donde $\overline{y}$ es el vector de medias muéstrales de los datos $Y_1, Y_2 \ldots, Y_m$ y S^{-1} es la inversa de la matriz de covarianza muestral.

Una vez calculadas las profundidades, se ordenan en forma ascendente, $Y_{[j]}$ denotará el punto muestral asociado con el j-ésimo valor de la profundidad más pequeña, entonces $Y_{[1]}$, $Y_{[2]}$, …,$Y_{[m]}$, serán los estadísticos de orden de $Y, Y_{[m]}$. Será el punto más "centrado" o profundo, mientras que $Y_{[1]}$ será el dato "más alejado" con respecto a los demás puntos, bajo la distribución H.

Una vez se tengan los datos muéstrales $X_i's$, se les calculará la profundidad aplicando la ecuación (1) o (2), utilizando para ello el vector de medias y la matriz de covarianzas de H_m y se denotará estas profundidades por $MD_{Hm}(x)$.

Para cada nuevo vector de observaciones de las características de calidad que provengan de la distribución , se calcula el estadístico de clasificación por rangos, r (•) mediante:

$$r(x_i) = \frac{\#\{y_j \setminus MD_{H_m}(y_j) \leq MD_{H_m}(x_i), j = 1,2, \ldots, m\ \}}{m + 1}$$

donde $r(x_i)$ es el i-ésimo estadístico de clasificación estimado correspondiente al i-ésimo vector de observaciones de la distribución G, y_j es el j-ésimo valor del estadístico de orden que se tiene cuando la $MD_{Hm}(y_j)$ es menor o igual a $MD_{Hm}(x_i)$ y denotará el número de observaciones de la distribución de referencia H .

La gráfica de comparación de rangos, se obtiene al graficar cada $r(x_i)$ contra el tiempo, con una línea central $CL=0,5$ y un límite de control inferior $LCI=\alpha$ siendo la probabilidad de cometer error tipo I. Si algún $r(x_i)$ se gráfica por debajo del LCI, entonces esta observación se declarará fuera de control, ya que este estadístico indica que tan atípico es x con respecto a la nube de puntos de los $Y_i's$. Liu (1995), demostró que el estadístico de clasificación por rangos sigue una distribución uniforme continua en el intervalo [0, 1].

La gráfica de comparación de rangos corresponde a una prueba de nivel de las hipótesis:

Ho: La distribución de referencia es igual a la distribución de monitoreo .
Versus
Ha: Existe un cambio en la ubicación o dispersión de la distribución H y G con respecto a la distribución de referencia .

2.4. Gráficos de Control Multivariados con Estimadores Alternativos

En la construcción del gráfico T^2 durante la fase I, se estiman los parámetros μ y $\sum$. Usualmente se hace mediante el vector de medias muéstrales y la matriz de covarianzas muestral, como se indicó en secciones anteriores. Estos estimadores son muy sensibles a desvíos de la normalidad y a la presencia de puntos atípicos, lo cual ha llevado al uso de procedimientos alternativos en la estimación de parámetros en control de procesos. Por lo tanto, es posible usar otros estimadores para estos parámetros y mejorar la potencia del gráfico.

Por ejemplo, Sullivan y Woodall (1996) recomiendan estimar $\sum$ mediante el vector de diferencias entre observaciones sucesivas, cuando se trabaja con observaciones individuales. El estimador de $\sum$ propuesto es

$$S^{\star} = \frac{1}{2(m-1)} \sum_{i=1}^{m-1} v_i v'_i$$

donde $v_i = X_{i+1} - X_i, \qquad i = 1,2,\ldots,m-1.$

y, el estimador de μ es el usual.

Sullivan y Woodall (1996) demostraron que el gráfico T^2 construido con base en la media muestral y S^* denotada como T^2_{sw} , detectaba en forma más rápida cambios en la media del proceso que el gráfico T^2 usual; sin embargo, no es efectiva cuando existen varios puntos atípicos en el proceso. Ante la presencia de grupos de puntos atípicos se ha recomendado el uso de estimadores robustos tanto para el caso univariado como el multivariado. Vargas (2003) propone construir el gráfico utilizando estimadores robustos de μ y $\sum$.

Ante la presencia de grupos de puntos atípicos, Vargas (2003) muestra que tanto el gráfico T^2 usual como el gráfico T^2_{sw} son ineficientes en su detección, debido al fenómeno de enmascaramiento. Propone entonces usar un gráfico T^2 con estimadores robustos de tal forma que este tipo de observaciones puedan ser identificados. Mediante simulaciones, Vargas (2003) muestra que el gráfico T^2 que utiliza estimadores de elipsoide de volumen mínimo (MVE) (Rousseeuw y van Zomeren (1990)), y estimadores de covarianza de determinante mínimo (MCD) (Rousseeuw y van Driessen (1999)) es eficiente ante la presencia de datos contaminados. Yañez y Cols. (2003) construyen gráficos T^2 con estimadores MVE y estimadores S (derivación de los MCD) y muestran también su eficiencia ante la presencia de varios puntos atípicos. Otro estimador robusto derivado del MCD conocido como estimadores de covarianza de determinante mínimo reponderado (RMCD), ha sido empleado en la construcción de gráficos T^2 .

Chenouri y Cols. (2009) concluyen que el grafico T2 RMCD es más eficiente que el gráfico usual, removiendo o no los puntos atípicos en la fase I. Además, presenta un mejor funcionamiento en la detección de puntos atípicos en la fase I que los gráficos T^2_{MCD} y T^2_{MVE} estudiados por Vargas (2003) y Jensen y Cols. (2007).

La distribución exacta con estimadores robustos basados en MVE y MCD no es posible. Así, se hace necesario obtener empíricamente, los límites de control para la fase I. Vargas (2003) y Jensen y Cols. (2007) estiman

los límites de control para gráficos T^2 robustos para la fase I, basado en simulaciones.

Para la construcción de la gráfica de comparación de rangos, las profundidades se calculan con los estimadores usuales de μ y ∑, sin considerar otro tipo de estimadores.

3. ESTUDIO DE SIMULACIÓN

3.1. Metodología

En esta sección se presentan los algoritmos utilizados para obtener el cálculo del ARL de los dos gráficos bajo estudio, iniciando con el cálculo de los límites de control superior del gráfico T^2 usando los diversos estimadores de μ y ∑.

Algoritmo 1. Cálculo de la distribución empírica de la estadística T^2 .

1. Se genera una muestra de tamaño $m=30$ de una normal p - variada $(p = 2)$ $X_l \sim N(\mu_0, \sum), l = 1,2,\ldots m$ se toma $\mu_0 = \mathbf{0}_{px1}$ y ∑, como aparece en la siguiente tabla.

Tabla 1. Diferentes matrices de covarianzas para $(p = 2)$ tomadas en la fase I y II

$p=2$	$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.2 \\ 0.2 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}$

2. Se estiman el vector de medias y la matriz de covarianzas con las observaciones X_l , usando los diferentes métodos para estimar los parámetros μ_0 y ∑, tales como el USUAL, SW, MVE, MCD y RMCD.

3. Para cada método, se genera una nueva observación $X_f \sim (\mu_0, \sum)$, que se considera de la fase II. Con esta observación y las observaciones obtenidas en 2. Se calcula el estadístico $T_f^2 = (X_f - \hat{\mu})' \hat{\Sigma}^{-1} (X_f - \hat{\mu})$.

4. Se repiten los pasos 1 al 3, 10000 veces. Se obtienen 10000 valores de la estadística T^2 .

5. De la distribución empírica de los T^2 se calcula el cuantil $(1 - \alpha)100\%$, el cual se considera el valor "tabulado" para con este determinar la significancia de la estadística T^2 calculada.

Algoritmo 2. Cálculo del ARL del gráfico de control de rangos.

1. Se generan $X_l \sim N(\mu_0, \sum), l = 1,2,\ldots.\mathrm{m}$ con $\mu_0 = \mathbf{0}_{\mathrm{px1}}$ y $\sum$, una matriz de las dadas en la tabla 1. En el caso de considerar outliers en este conjunto de datos, se genera $X_s \sim N_p(\mu, \sum), \mathrm{s} = 7, 24, 21 \ \mathrm{con}\ \mu_0 = \mathbf{0,5}_{\mathrm{px1}}$. Se estima μ_0 y $\sum$ con un método (USUAL, SW, MVE, MCD y RMCD).

2. Se calcula las profundidades de Mahalanobis

$$MD(X_l) = \frac{1}{1 + (X_f - \hat{\mu})' \hat{\Sigma}^{-1} (X_f - \hat{\mu})}$$

para las m observaciones y se ordenan de menor a mayor.

3. Se genera una nueva observación $X_f \sim N_p(\mu_1, \sum)$ que se considera de la fase II, donde $\mu_1 = v_{\mathrm{px1}}$ con $v = 0, 0.5, 1.0, 1.5, 2.0, 2.5$ y 3.0 (cambio en media).

4. Se calcula la profundidad de Mahalanobis de X_l y se comparan con las profundidades de las m observaciones de la fase I. Se calcula la estadística.

$$\gamma(X_f) = \frac{\#\{X_l / MD(X_l) \leq MD(X_f),\ \ l = 1,2,\cdots,m\}}{(m+1)}$$

5. Se genera una nueva observación X_f y se repite el paso 4, hasta que $\gamma \leq 0.05$. Se cuentan los puntos hasta que se produjo la señal. Este conteo se toma como la longitud de corrida (RL) del gráfico.

6. Se repiten los puntos 1 al 5, 5000 veces. Se obtienen 5000 RL, los cuales se promedian y se calcula la longitud promedio de corrida (ARL) del gráfico de rangos.

Algoritmo 3. Cálculo del ARL del gráfico de control T^2.

1. Se realizan los puntos 1. y 3. del algoritmo 2.
2. Se calcula la estadística T^2 a X_f. Si,

$$T^2_{X_f} = (X_f - \hat{\mu}_0)' \hat{\Sigma}^{-1} (X_f - \hat{\mu}_0) < T^2_{Tab},$$

el punto no produce señal. (T^2_{Tab} , es el obtenido con el algoritmo 1).

3. Se genera una nueva observación X_f y se repite el paso 2, hasta que .

$$: T^2_{X_f} \geq T^2_{Tab}.$$

4. Se repiten los puntos 1 al 3, 5000 veces. Se promedia los 5000 RL para obtener el ARL del gráfico T^2 .

3.2. Resultados

En esta sección se relaciona un conjunto de figuras, las cuales están organizadas de la siguiente forma: En la tabla superior se ubican los límites de control para el gráfico T^2 para diferentes estimadores de μ y ∑, obtenidos con el algoritmo 1 de la sección anterior. Seguidamente, se encuentran los cálculos correspondientes del ARL de los gráficos T^2 y de rangos con y sin outliers respectivamente, utilizando los algoritmos 2 y 3 de la sección anterior.

Límite\Estimador	USUAL	S&W	MVE	MCD	RMCD
T^2	6.833094	7.322023	13.16408	12.40650	7.750732

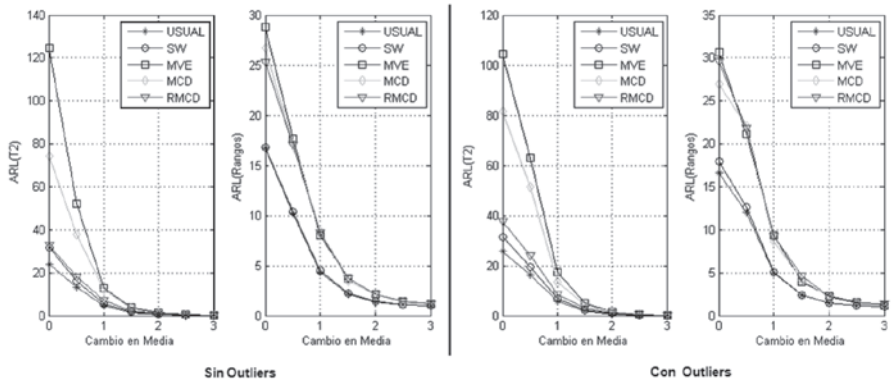

Figura 1. Estimación de ARL con los métodos USUAL, SW, MVE, MCD Y RMCD para los gráficos y de Rangos, sin presencia y con presencia de outliers, con $Cov(X_i, X_j) = 0$

Límite\Estimador	USUAL	S&W	MVE	MCD	RMCD
T^2	7.379735	7.737582	13.90578	13.36145	8.496861

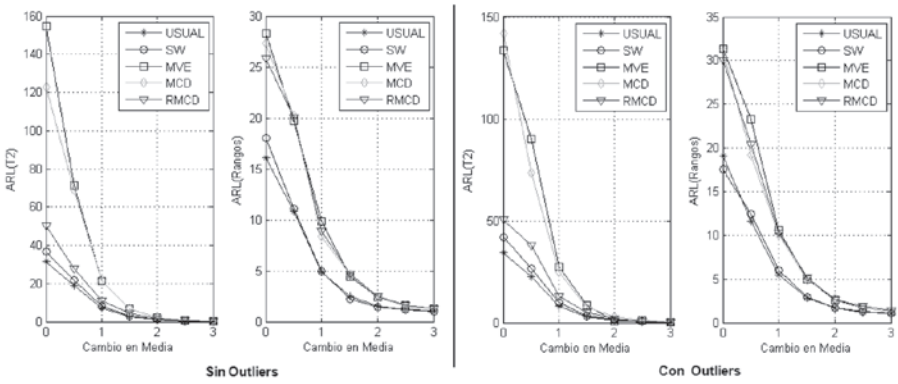

Figura 2. Estimación de ARL con los métodos USUAL, SW, MVE, MCD Y RMCD para los gráficos y de Rangos, sin presencia y con presencia de outliers, con $Cov(X_i, X_j) = 0.2$

Límite\Estimador	USUAL	S&W	MVE	MCD	RMCD
T^2	7.10172	7.57005	14.03958	13.28154	8.356006

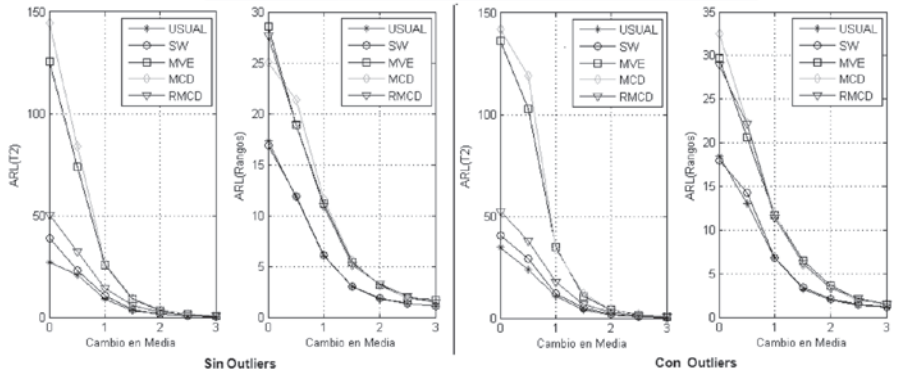

Figura 3. Estimación de ARL con los métodos USUAL, SW, MVE, MCD Y RMCD para los gráficos y de Rangos, sin presencia y con presencia de outliers, con $Cov(X_i, X_j) = 0.5$

Límite\Estimador	USUAL	S&W	MVE	MCD	RMCD
T^2	7.092119	7.568998	13.86348	13.36501	8.463319

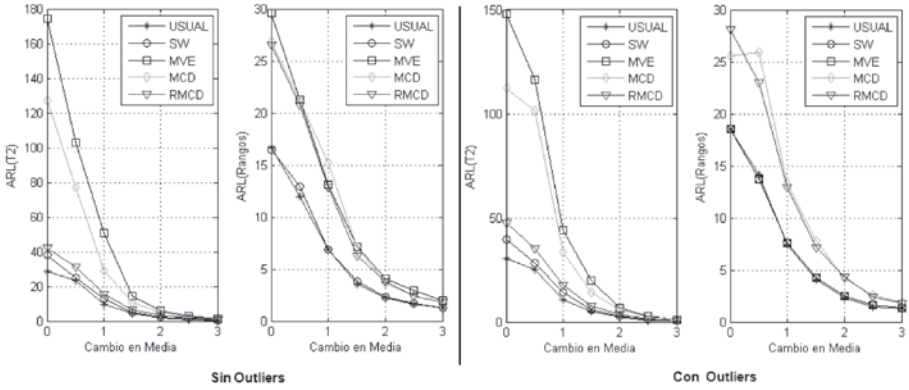

Figura 4. Estimación de ARL con los métodos USUAL, SW, MVE, MCD Y RMCD para los gráficos y de Rangos, sin presencia y con presencia de outliers, con $Cov(X_i, X_j) = 0.9$

El valor $ARL_0 = \frac{1}{\alpha} = \frac{1}{0.05} = 20$ teórico es el tomado para la construcción de los gráficos.

Tabla 2. Diferentes matrices de covarianzas para tomadas en la fase I y II

$p = 3$	$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.2 & 0 \\ 0.2 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.5 & 0 \\ 0.5 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.9 & 0 \\ 0.9 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	$\begin{pmatrix} 1 & 0.2 & 0.5 \\ 0.2 & 1 & 0.9 \\ 0.5 & 0.9 & 1 \end{pmatrix}$

Resultados similares a los obtenidos para los casos de se encontraron para $p = 2$ se encontraron para $p = 3$.

3.3. Interpretaciones

1) El gráfico de rangos presenta un ARL menor que el gráfico T^2, usando cualquier estimador de μ y $\sum$. El gráfico de Rangos presenta valores ARL próximos al teórico ($\delta = 0$) en especial con los estimadores USUAL y SW con presencia o no de outliers. En el gráfico T^2 cuando se usan los estimadores MVE y MCD pierde eficiencia en detección de señal, con presencia o no de outliers.

2 La mayor reducción del ARL se produce cuando el descentrado es de $\delta = 1.5$ y $\delta = 2.0$ en el gráfico de rangos y ocurre una reducción aún mayor en el caso del gráfico T^2.

3) Ante cambios pequeños en media $\delta = 0.5$ y $\delta = 1.0$ la que presenta una mayor eficiencia en detección de cambio en el vector de medias del proceso es el gráfico T^2.

4) Para cualquier grado de asociación entre las variables (nula, baja, media o alta) no afecta el funcionamiento de los dos gráficos. Igual ocurre con la presencia o no de outliers en este caso considerado el 10% de las observaciones del conjunto de datos históricos como datos atípicos.

4. CONCLUSIONES

En este artículo se presentó la evaluación de la eficiencia de los gráficos de control y de rangos, a través de un estudio de simulación de datos bivariados, cuando se produce cambio en el vector de medias del proceso ante diversas correlaciones entre las variables y presencia de datos atípicos. La eficiencia fue medida a través del cálculo de ARL de los gráficos bajo control estadístico, llegando a determinarse que el gráfico de rangos es mejor en la detección de señal ante cualquier estimador usado, en especial el USUAL y SW en comparación con el gráfico T^2. Aunque para detectar cambios pequeños en media es mejor el gráfico T^2 .

Se recomienda para estudios posteriores trabajar con datos no normales y contaminaciones de los datos históricos más severos, en cantidad de datos atípicos y en el desvío de la media del proceso para comparar la eficiencia de los dos gráficos de control.

5. AGRADECIMIENTOS

Los autores agradecen el apoyo al comité central de investigaciones de la universidad del Tolima. Este trabajo es un producto derivado del proyecto de investigación número 200110.

BIBLIOGRAFÍA

- Alt, F.B y Smith, N.D. (1988). ***Multivariate Process Control, in P.R.*** Krishnaiah y C.R. Rao, eds, Handbook of Statistics, 7, North holland, Amsterdam.

- Chenouri, S., Steiner, S.H. y Variant, A.M. (2009). ***A Multivariate***

Robust Control Chart for Individual. Journal of Quality Technology 41, No: 259-271.

- Crosier, R.B. (1988). ***"Multivariate Generalizations of Cumulative Sum Quality Control chemes".*** Technometrics, 30, 291-303.

- Hotelling, H. (1947). ***Multivariate Quality Control. Techniques of Statistical Analysis, S.C*** Eisenhart, M.W. Hastay, y W.A. Wallis. New York: McGraw Hill.

- Jensen, W.A., Birch, J.B. y Woodall, W.H. (2007). ***High Breakdown Estimation Methods for Phase I Multivariate Control Charts.*** Quality and Reliability Engineering International, 23(5), 1638-1665.

- Liu, R. (1995). ***Control Chart for Multivariate Processes. Journal American Statistics Association.*** Theory and Methods. 90, No 432: 1380-1387.

- Lowry, C.A.; Woodall, W.H.; Champ, C.W.; y Rigdon, S.E. (1992). ***A Multivariate Exponentially Weighted Moving Average Control Chart.*** Tecnometrics, 34, pp.46-53.

- Lowry, C.A.; y Montgomery, D.C. (1995). ***A Review of Multivariate Control Charts.*** IIE Transactions, 27, 800-810.

- Pignatiello, J.J. Jr.; y Runger, G.C. (1990). ***Comparisons of Multivariate CUSUM Charts"***. Journal of Quality Technology, 22, 173-186.

- Rousseeuw, P.J. y van Driessen, K. (1999). ***A Fast Algorithm for the Minimum Covariance Determinant Estimator.*** Technometrics, 41, 212-223.

- Rousseeuw, P.J. y van Zomeren, B.C. (1990). ***Unmasking Multivariate Outliers and Leverage Points (with discussion).*** Journal of the American Statistical Association, 85, 633-651.

- Sullivan, J.H., Woodall, W.H. (1996). ***A Comparison of Multivariate Control Charts for Individual Observations.*** Journal of Quality Technology, 28: 398-408.

- Tracy, N.D., Young, J.C. y Mason, R.L. (1992). ***Multivariate Control Charts for Individual Observations.*** Journal of Quality Technology, 24, 88-95.

- Vargas, J.A. (2003). ***Robust Estimation in Multivariate Control Charts for Individual Observations.*** Journal of Quality Technology, 35(4), 367-376.

- Velásquez, H. y Moreno, D.P. (2009). ***Tratamientos a Puntos Atípicos Identificados por Gráficos de Control Multivariados.*** XIX Simposio Nacional de Estadística, Medellin Colombia. Comunicación.

- Wierda, S.J. (1994). ***Multivariate Statistical Process Control-Recent. Results and Directions for Future Research***. Statistica Neerlandica, 48,147-168.

- Willems, G, Pison G., Rousseeuw, P.J. y Van Aelst, S. (2002). ***A Robust Hotelling Test***. Metrika 55:125-138.

- Yañez, S. Vargas, J.A., y González, N. (2003). ***Carta T2 con Base en Estimadores Robustos de los Parámetros.*** Revista Colombiana de Estadística, 26, 159-179.

- Zertuchi, F. y Cantú, M. (2008). ***Una Comparación del Desempeño de las Cartas de Control T2 de Hotelling y de Clasificación por Rangos.*** Ingeniería, Investigación y Tecnología, 9, Nº 3, 205-215.

- Zuo, Y. y Serfling, R. (2000). ***Structural Properties and Converge Results for Contours of Sample Statistical Depth Functions.*** The Annals of Statistics, 483-499.