

Evaluación de la efectividad del método de Murphy para la interpretación de señales en el gráfico de control multivariado T^2

Alex J. Zambrano^I, Hector F. López^{II}

Resumen. Frecuentemente puede ser problemático identificar cuando una característica o grupo de características de calidad presenta un proceso fuera de control en el control multivariado de procesos estadísticos, ya que la calidad es representada por el control simultáneo de varias variables aleatorias correlacionadas.

En este trabajo se estudia la metodología de Murphy para detectar las causas asignables en una señal de fuera de control en el gráfico de control multivariado T^2 . Se propone evaluar la efectividad de este método, utilizando simulación de procesos multivariados, cuando ocurren cambios en el vector de medias, e igualmente cuando las variables se encuentran correlacionadas. Se presenta un gráfico que muestra los resultados obtenidos.

Palabras clave: Control de calidad multivariado, estadístico T^2 de Hotelling, gráfico de control T^2 , selección de variables, análisis discriminante.

Abstract. To identify when a characteristic or group of characteristics of quality presents a process out of control in the multivariate statistical process control can be a problem, because the quality is represented by the simultaneous control of correlated random variables.

In this paper we study the methodology of Murphy to detect assignable causes in an out of control signal in T^2 control chart. The purpose is to evaluate the effectiveness of this method, using multivariate process simulation, when occur changes in the mean vector, and when the variables are correlated. A graph shown the results obtained.

Key words: Multivariate quality control, Hotelling's T^2 statistic, T^2 control chart, selection of variables, discriminant analysis.

I Estudiante de Maestría en Estadística, Universidad Nacional de Colombia, Bogota, Colombia, aj-zambranoc@unal.edu.co.

II Profesional en Matemáticas con énfasis en Estadística, Universidad del Tolima, Ibagué, Colombia, heticor86@hotmail.com.

1. INTRODUCCIÓN

Según Bersimis, Psarakis, y Panaretos (2007) los gráficos de control multivariados son una herramienta poderosa en el Control de Proceso Estadístico en cuanto a la identificación en un proceso fuera de control. Woodall y Montgomery (1999) enfatizan la necesidad de investigar más en esta área dado que la mayoría de procesos involucra un gran número de variables que son correlacionadas. Jackson (1991) comenta que cualquier procedimiento en control de calidad multivariado debe cumplir las siguientes condiciones:

1. Responder la pregunta “¿Esta el proceso en control?”.
2. La probabilidad conjunta para el error tipo I debe especificarse.
3. Las relaciones entre las variables deben tenerse en cuenta.
4. Se debe dar un procedimiento que permita responder la pregunta “Si el proceso está fuera de control, ¿Cuál es el problema?”.

La última pregunta ha provisto un interesante tema para muchos investigadores en los últimos años. Woodall y Montgomery (1999) manifiestan que aunque hay dificultad en la interpretación de señales en los gráficos de control multivariado se ha trabajado en métodos de reducción y técnicas gráficas. Maravelakis (2003) describe algunos de estos trabajos, entre ellos destacamos los realizados por Alt (1985) quien trabaja usando p gráficos de control univariado con límites Bonferroni; Jackson (1991) propone usar la región de control elíptica y también componentes principales, y Mason, Tracy, y Young (1995, 1997) usan la descomposición del estadístico T^2 en partes independientes (llamada descomposición MYT) las cuales reflejan la contribución de las variables de manera individual y condicional. De este último trabajo algunas propuestas por otros autores incluyen esta descomposición, entre ellos Roy (1958), Murphy (1987), Doganaksoy, Faltin, y Tucker (1991), Hawkins (1991, 1993), Timm (1996) y Runger, Alt, y Montgomery (1996). La metodología de Murphy (1987) consiste en utilizar el Análisis Discriminante con el cual se detecta la variable o el conjunto de variables que hacen que el proceso presente una señal fuera de control. En trabajos anteriores se evaluó la efectividad de la descomposición MYT Zambrano y Zambrano (2008), para este se pretende evaluar la metodología de Murphy (1987) para la interpretación de señales.

Este artículo pretende abordar los conceptos básicos del gráfico de control multivariado T^2 para una observación individual (sección 2), consecuentemente se

describe la metodología propuesta por Murphy (1987) (sección 3), el procedimiento para evaluar esta metodología (sección 4), los resultados obtenidos (sección 5) y las conclusiones de los mismos (sección 6).

2. GRÁFICO DE CONTROL T^2

Este gráfico propuesto por Hotelling (1947) permite monitorear la distancia estadística entre un vector de observaciones y el vector de medias, para los casos en que se utilizan observaciones individuales. Suponiendo que los datos se distribuyen normal p variada con vector de medias $\boldsymbol{\mu}_0$ y matriz de covarianzas $\boldsymbol{\Sigma}$ conocidos, la estadística T^2 para la i -ésima observación individual $\mathbf{X}_i^t = (X_{i1}, \dots, X_{ip})$ viene dada por

$$T^2(\mathbf{X}_i) = (\mathbf{X}_i - \boldsymbol{\mu}_0)^t \boldsymbol{\Sigma}^{-1} (\mathbf{X}_i - \boldsymbol{\mu}_0) \quad i = 1, 2, \dots, n. \quad (1)$$

Si se desconocen los parámetros de la normal p variada, se utilizan sus estimaciones $\bar{\mathbf{X}}$ y \mathbf{S} obtenidas a partir de datos históricos, por lo cual la ecuación (1) queda expresada como

$$T^2(\mathbf{X}_i) = (\mathbf{X}_i - \bar{\mathbf{X}})^t \mathbf{S}^{-1} (\mathbf{X}_i - \bar{\mathbf{X}}) \quad i = 1, 2, \dots, n, \quad (2)$$

donde el vector de medias muestrales viene dada por

$$\bar{\mathbf{X}} = \frac{1}{n} \sum_{i=1}^n \mathbf{X}_i,$$

y la matriz de covarianzas muestrales

$$\mathbf{S} = \frac{1}{n-1} (\mathbf{X}_i - \bar{\mathbf{X}})(\mathbf{X}_i - \bar{\mathbf{X}})^t.$$

El gráfico de control consiste en graficar la estadística T^2 v.s. el número de observaciones. Sin embargo, dependiendo de si se conocen o no los parámetros de la distribución normal p variada, se considera una distribución diferente a un nivel de significancia α Díaz (2002). Está distribución, según el caso, permite asociar un límite de control superior el cual se representa en el gráfico de control como una línea recta con la cual se puede indicar si un proceso está bajo control. Está es una diferencia notable frente a los gráficos de control univariados ya que no tienen importancia el límite central ni el de control inferior debido a que la estadística T^2 por ser cuadrática nunca es negativa. Además, no presenta ningún

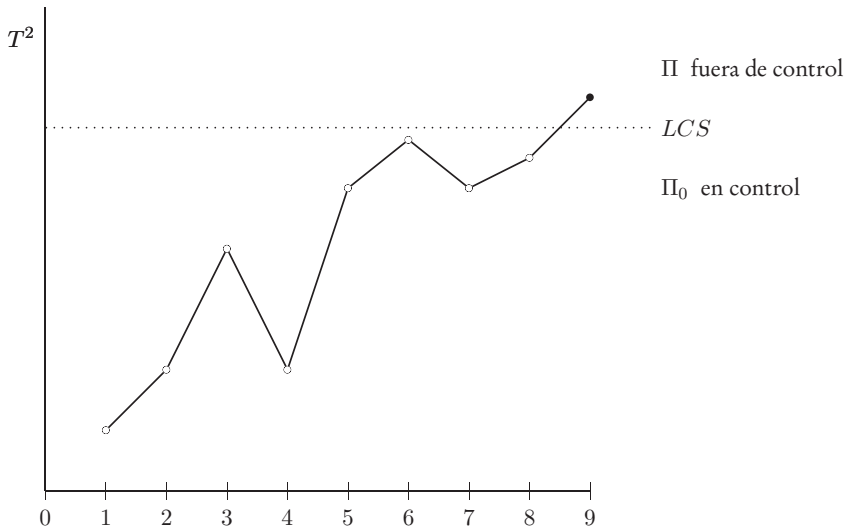


Figura 1: Gráfico de control T^2 con límite de control superior LCS .

problema si es muy cercano a cero (indicaría que la observación es muy cercana al vector de medias de las variables) (Figura 1).

Si la variabilidad de las observaciones en el proceso es natural o común, se presenta un gráfico de control donde cada estadística está por debajo del límite de control superior, se dice entonces que el proceso está bajo control. En caso contrario se dice que el proceso está fuera de control, y se deberá a que la variabilidad de las observaciones en el proceso no es natural o presenta alguna causa asignable o especial.

El gráfico de control T^2 se aplica en dos fases que se conocen como fase I y II; en la primera de ellas se toma un conjunto de datos históricos y se construye el gráfico de control T^2 para establecer si el proceso se encontraba bajo control cuando se sacaron n observaciones individuales preliminares y se calcularon los estadísticos \bar{X} y S . El objetivo en la fase I es obtener un conjunto de observaciones homogéneas bajo control a fin de establecer el límite de control superior para la fase II, el cual consiste en el monitoreo de la producción futura. En ocasiones se le denomina análisis retrospectivo Montgomery (2004).

2.1 Límites de control para la fase I

La fase I consta de dos etapas. La etapa 1 consiste en construir a partir de un conjunto de datos histórico el límite de control superior, para determinar si el proceso está bajo control. En esta etapa la distribución de la estadística T^2 dada por la ecuación (2) se distribuye

$$T^2 \sim \frac{(n-1)^2}{n} B\left(\frac{p}{2}, \frac{n-p-1}{2}, \alpha\right).$$

Para determinar si el proceso se encuentra bajo control se establece el límite de control superior asociado a la distribución del estadística anterior dada por

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

donde $B_{(\delta_1, \delta_2, \alpha)}$ es el percentil $(1 - \alpha)$ de la distribución beta con parámetros δ_1 y δ_2 . Cuando una observación hace que el proceso se encuentre fuera de control a está se la conoce como señal. Si el gráfico T^2 indica que hay una señal, se debe llevar a cabo una investigación con el objeto de encontrar causas especiales que la hubieran producido. Las observaciones multivariadas que correspondan a causas especiales se eliminan. Entonces se calculan de nuevo el límite de control superior para un examen retrospectivo basado en las observaciones restantes y el procedimiento se repite.

En la etapa 2 se prueba si el proceso permanece bajo control cuando nuevas observaciones sean seleccionadas. La estadística T^2 sigue una distribución F , por lo cual el gráfico de control T^2 utiliza en esta etapa un límite de control superior dado por,

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

donde n es el tamaño del observaciones en el conjunto de datos históricos con las que finalmente se estimaron los parámetros en la etapa 1 y $F_{(p, m-p, \alpha)}$ es el percentil $(1 - \alpha)$ de la distribución F con parámetros p y $n - p$. Una vez estabilizado el proceso en la etapa 2, se asume que los estimadores finales \bar{X} y S son los verdaderos valores de los parámetros bajo control Vargas (2006).

2.2 Límites de control para la fase II

Para la fase II la carta T^2 , dado que se conocen los parámetros la estadística T^2 se construyen como la ecuación (1), en este punto el límite de control superior

viene dado por

$$LCS = \chi^2_{(p,\alpha)}, \tag{3}$$

donde $\chi^2_{(p,\alpha)}$ es el percentil $(1 - \alpha)$ de la distribución χ^2 central con parámetro p .

Un desarrollo de los límites cuando se consideran subgrupos de observaciones mayores que uno se realiza de manera similar en Vargas (2006). En adelante se asumirá que el proceso ha pasado la Fase I, por lo que el proceso se encuentra bajo control.

3. METODOLOGÍA DE MURPHY

Este procedimiento es propuesto por Murphy (1987), subcaso de la descomposición de MYT según Maravelakis (2003). Se asume que el proceso está en fase II descrito anteriormente (ver sección 2), por lo cual cada observación individual \mathbf{X}_i se distribuye normal p variada con vector de medias $\boldsymbol{\mu}_0$ y matriz de covarianza $\boldsymbol{\Sigma}$. Estas ideas desarrolladas son extendibles para el caso donde $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$ puedan ser estimados Murphy (1987).

La población bajo control se denotará por $\Pi_0 \sim N_p(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$. La familia de poblaciones fuera de control se denotara por $\Pi \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma}), \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. El procedimiento hace monitoreos cuando hay algún cambio en la media $\boldsymbol{\mu}_0$.

Para tomar la decisión sobre i el proceso está o no bajo control, se utiliza la estadístico T^2 para una observación individual \mathbf{X}_i dado por la ecuación (1) la cuál es equivalente a verificar la hipótesis $H_0 : \boldsymbol{\mu} = \boldsymbol{\mu}_0$ v.s. $H_1 : \boldsymbol{\mu} \neq \boldsymbol{\mu}_0$. Seleccionando el límite de control superior (LCS) correspondiente a la ecuación (3), se define la siguiente regla de decisión

Si $T^2 \leq LCS$; entonces $\mathbf{X}_i \in \Pi_0$ (proceso bajo control).

Si $T^2 > LCS$; entonces $\mathbf{X}_i \notin \Pi_0$ (proceso fuera de control).

3.1 Ventajas de utilizar el gráfico de control T^2

La ventaja principal de utilizar estadístico T^2 en el gráfico de control T^2 , es la propia reflexión sobre la estructura de correlación de las poblaciones (están implícitas las correlaciones de las poblaciones, ya que $\boldsymbol{\Sigma}$ es conocida), aunque también se presenta la misma situación con un buen estimador de $\boldsymbol{\Sigma}$. En la práctica, el no tener en cuenta la correlación entre las variables es una desventaja puesto se deben utilizar cartas independientes para monitorear cada una de las p características de calidad de manera individual. Haciendo esto se hacen pruebas

simultáneas sobre cartas individuales permitiendo que cada error de tipo I sea α/p , en lugar de α , lo cual es una dificultad.

Otra razón para utilizar el gráfico de control T^2 es la facilidad de cálculo y de construcción simple del mismo ya que sólo requiere comparar valores T^2 con el límite de control superior (Figura 1).

3.2 Interpretación de una señal y selección de la variable causante

Se propone utilizar el Análisis Discriminante Díaz (2002). Se plantea interpretar una señal fuera de control y una prueba sobre cuál de las variables (o subconjunto de ellas) causaron la señal.

Dada una observación \mathbf{X}_* que produce una señal en el gráfico de control T^2 ($T^2(\mathbf{X}_*) > LCS$), la pregunta de interés inmediato es: ¿Cuál de las p variables, o subconjunto p_1 de ellas ($p = p_1 + p_2$) causaron la señal? Murphy (1987) señala que particionando la observación $\mathbf{X}_*^t = [\mathbf{X}_*^{(1)}, \mathbf{X}_*^{(2)}]$, donde $\mathbf{X}_*^{(1)}$ es el subconjunto de las p_1 variables las cuales se sospecha causaron la señal, y $\mathbf{X}_*^{(2)}$ es el conjunto de las p_2 variables restantes.

Además, denotando con T_p^2 la distancia cuadrada completa,

$$T_p^2 = T^2(\mathbf{X}_*) = (\mathbf{X}_* - \boldsymbol{\mu}_0)^t \boldsymbol{\Sigma}^{-1} (\mathbf{X}_* - \boldsymbol{\mu}_0)$$

y con $T_{p_1}^2$ la distancia reducida correspondiente al subconjunto p_1 ,

$$T_{p_1}^2 = T^2(\mathbf{X}_*^{(1)}) = (\mathbf{X}_*^{(1)} - \boldsymbol{\mu}_0^{(1)})^t \boldsymbol{\Sigma}_{11}^{-1} (\mathbf{X}_*^{(1)} - \boldsymbol{\mu}_0^{(1)})$$

donde $\boldsymbol{\mu}_0$ y $\boldsymbol{\Sigma}$ están particionadas como lo esta \mathbf{X}_* . Finalmente, la siguiente diferencia es calculada

$$D = T_p^2 - T_{p_1}^2, \tag{4}$$

el cual $D \sim \chi_{p_1}^2$. Si D es grande, se rechaza la hipótesis nula de que el subconjunto p_1 causó la señal. Si es pequeño, no se rechaza. Pruebas similares son bien conocidas en el análisis discriminante y en el análisis de regresión para tratar con problemas de selección de variables Díaz (2002). Si $(\boldsymbol{\mu}_0, \boldsymbol{\Sigma})$, son estimados, la prueba apropiada para D es una distribución F Seber (1984).

3.3 Algoritmo de Murphy para seleccionar variables fuera de control

El siguiente algoritmo describe y reduce el cálculo computacional para considerar señales fuera de control. Se consideran observaciones individuales. Si se desea trabajar con observaciones en subgrupos mayores que uno, este mismo algoritmo se describe en Murphy (1987).

Paso 1 Llevar a cabo una prueba T^2 con un nivel de significancia α . Si la condición se encuentra fuera de control entonces se continua con el paso 2.

Paso 2 Calcular los p individual $T_1^2(X_i)$ equivalente a mirar las p cartas individuales y calcular las p diferencias $D_{p-1}(i) = T_p^2 - T_1^2(X_i)$. Seleccionar el $\min\{D_{p-1}(i)\} = D_{p-1}(r)$ y probar esta mínima diferencia. Si $D_{p-1}(r)$ es no significativa entonces la variable r -ésima es la única que requiere atención. Si $D_{p-1}(r)$ es significativa continúe con el paso 3.

Paso 3 Calcular las $p - 1$ diferencias $D_{p-2}(r, j) = T_p^2 - T_2^2(X_r, X_j)$, $1 \leq r, j \leq p$ y $r \neq j$. Seleccionar el $\min\{D_{p-2}(r, j)\} = D_{p-2}(r, s)$ y pruebe la mínima diferencia. Si $D_{p-2}(r, s)$ no es significativa entonces las variables r -ésima y s -ésima son las únicas que requieren atención. Si $D_{p-2}(r, s)$ es significativa entonces continúe con el paso 4.

Paso 4 Similar al paso 3.

Paso . Similar al paso 3 y 4.

Paso p Si la última $D_{p-(p-1)}$ es significativa, entonces, todas las p variables requieren atención.

Murphy (1987) recomienda que en las pruebas $D_{p-i} \sim \chi_{(p-i, \alpha^*)}^2$ se use un nivel de significancia en el intervalo $0.1 \leq \alpha^* \leq 0.2$.

Ejemplo 3.1. Considérese un proceso de tres variables, el cual se distribuye multivariadamente normal con parámetros poblacionales dados por el vector de medias ceros ($\mu = \mathbf{0}$) y matriz de varianzas la idéntica de orden 3 ($\Sigma = \mathbf{I}_3$), para observaciones individuales. Se calculará el estadístico T^2 para la observación $\mathbf{X}_i^t = [-0.1736, 2.9356, 0.0379]$.

$$T^2(\mathbf{X}_i) = (\mathbf{X}_i - \mu)^t \Sigma^{-1} (\mathbf{X}_i - \mu) = 8.649026$$

Comparando este valor con el límite de control superior dado por la ecuación (3) para $\alpha = 0.05$ se observa que se presenta una señal ($LCS = \chi_{(3, 0.05)}^2 = 7.814728$).

Ahora según el algoritmo anterior calculan los términos individuales $T_1^2(X_i)$ ($i = 1, 2, 3$)

$$T_1^2(X_1) = 0.0314, \quad T_1^2(X_2) = 8.6177, \quad T_1^2(X_3) = 0.0014$$

Se calculan las diferencias utilizando la ecuación (4) y se elige la mínima diferencia de las tres anteriores ($D_2(2) = 0.0313$). Se determina que esta diferencia no

es significativo ($\chi^2_{(2,0.1)} = 4.60517$), entonces la variable X_2 requiere atención ya que es la causante de la señal fuera de control.

4. METODOLOGÍA

Para evaluar la efectividad de la metodología de Murphy (1987) se propone estudiar cambios en las componentes de vector de medias y correlación entre variables tal como lo describimos a continuación.

Se propone generar observaciones individuales de un proceso multivariado normal estándar con máximo tres variables. Se toma una muestra de $n = 100$ bajo control. Como se conocen los parámetros entonces se considera el límite de control superior dado por (3) con un nivel de significancia de $\alpha = 0.05$. Se hacen cambios a las componentes del vector de medias, o a la matriz de covarianzas de la siguiente manera:

1. Se cambia una o dos o las tres componentes del vector de medias, aumentando está con valores $\mu = 0.5, 1, 1.5, 2, 3, 5$.
2. Se realizaran cambios a la matriz de covarianzas para obtener correlaciones entre las variables X_1 y X_2 del proceso con los siguientes valores $\rho_{12} = 0.1, 0.5, 0.8$. Con esto se espera observar sus efectos.

Luego de encontrar el proceso fuera de control en un tiempo t se aplica la metodología de Murphy (1987) en este tiempo. Se interpretarán las causas que determinan la señal fuera de control. Esto se realizarán 5000 veces para evaluar la efectividad de esta descomposición, en términos de probabilidades, lo que facilitará la interpretación de la misma.

5. RESULTADOS

La tabla 1 resume los resultados obtenidos de las diferentes corridas realizadas. La primera columna representa las correlaciones para las variables (X_1, X_2) , la segunda columna representa los valores de los corrimientos en vector de medias. Las siguientes columnas representan las probabilidades en las que el método de Murphy detecta adecuadamente las variables a las cuales se les realizó el cambio; e.d., si la columna 4 de la fila 4 hay un valor de 0.65 indica que la probabilidad de que el método de Murphy detecte cuando se produce un corrimiento en la segunda componente del vector de medias de $\mu = 1.5$, cuando $\rho_{12} = 0$ es de un 65 %. A estos valores los llamaremos probabilidades de detención efectivas en las variables que producen señal. Recordemos que los corrimientos en

Tabla 1: Probabilidades de detección efectivas de las variables según corrimientos en el vector de medias y correlación entre la 1 y 2 variable

ρ_{12}	μ	$(\mu, 0, 0)$	$(0, \mu, 0)$	$(0, 0, \mu)$	$(\mu, \mu, 0)$	$(\mu, 0, \mu)$	$(0, \mu, \mu)$	(μ, μ, μ)
0	0	0.25	0.26	0.24	0.07	0.07	0.07	0.04
	0.5	0.35	0.35	0.34	0.08	0.09	0.08	0.04
	1	0.50	0.51	0.51	0.15	0.14	0.15	0.08
	1.5	0.63	0.65	0.64	0.22	0.23	0.23	0.17
	2	0.72	0.72	0.72	0.34	0.35	0.35	0.33
	3	0.82	0.82	0.81	0.61	0.62	0.60	0.72
	5	0.82	0.81	0.83	0.80	0.80	0.81	0.99
0.1	0	0.25	0.24	0.26	0.07	0.07	0.07	0.04
	0.5	0.35	0.34	0.35	0.10	0.09	0.09	0.05
	1	0.51	0.51	0.51	0.17	0.15	0.15	0.09
	1.5	0.62	0.64	0.65	0.25	0.22	0.24	0.17
	2	0.74	0.72	0.72	0.38	0.35	0.35	0.32
	3	0.81	0.82	0.82	0.63	0.61	0.62	0.71
	5	0.82	0.83	0.82	0.80	0.80	0.79	0.99
0.5	0	0.21	0.22	0.22	0.15	0.06	0.06	0.06
	0.5	0.30	0.32	0.30	0.21	0.08	0.08	0.07
	1	0.49	0.50	0.47	0.28	0.15	0.15	0.13
	1.5	0.63	0.64	0.62	0.37	0.24	0.23	0.22
	2	0.72	0.74	0.72	0.47	0.34	0.35	0.37
	3	0.82	0.82	0.81	0.67	0.63	0.63	0.73
	5	0.82	0.83	0.83	0.80	0.80	0.79	0.99
0.8	0	0.18	0.17	0.18	0.27	0.05	0.05	0.09
	0.5	0.27	0.27	0.27	0.32	0.07	0.08	0.11
	1	0.47	0.47	0.44	0.41	0.14	0.14	0.18
	1.5	0.62	0.63	0.58	0.51	0.23	0.21	0.28
	2	0.72	0.73	0.70	0.57	0.35	0.34	0.42
	3	0.84	0.82	0.80	0.73	0.63	0.63	0.76
	5	0.83	0.83	0.82	0.81	0.79	0.81	0.99

el vector de medias para que produzca una señal pueden ser corrimientos de manera individual $(\mu, 0, 0)$, $(0, \mu, 0)$, $(0, 0, \mu)$, corrimientos en dos componentes $(\mu, \mu, 0)$, $(\mu, 0, \mu)$, $(0, \mu, \mu)$ y corrimiento en las tres componentes (μ, μ, μ) . Nótese que cuando $\rho_{12} = 0$ y el corrimiento se hace en las tres componentes del vector de medias para $\mu = 0$ la probabilidad de detectar una observación fuera de control cuando no se hicieron corrimientos es alrededor del 4%, lo que se acerca bastante al nivel de significancia definido.

Realizando un gráfico de perfiles (Figura 2) se observa que las probabilidades de detección efectivas aumentan cuando el corrimiento en las medias es alto. Además se observa un comportamiento similar en los valores de probabilidad de detención en las corrimientos del vector de medias de manera individual,

independiente de la correlación ρ_{12} , por tanto no se afecta la efectividad de detección en las variables corridas de manera independiente. Así mismo cuando existe correlación baja o no la hay, se observan que los valores de probabilidad son similares (Figuras 2a y 2b). Por el contrario, cuando la correlación es moderada o grande, se observan un aumento en la probabilidad de detección cuando se realiza corrimientos simultáneos en las variables (X_1, X_2) (Figuras 2c y 2d). Cuando se tiene una correlación alta se afectara la probabilidad de detección efectividad (Figura 2d).

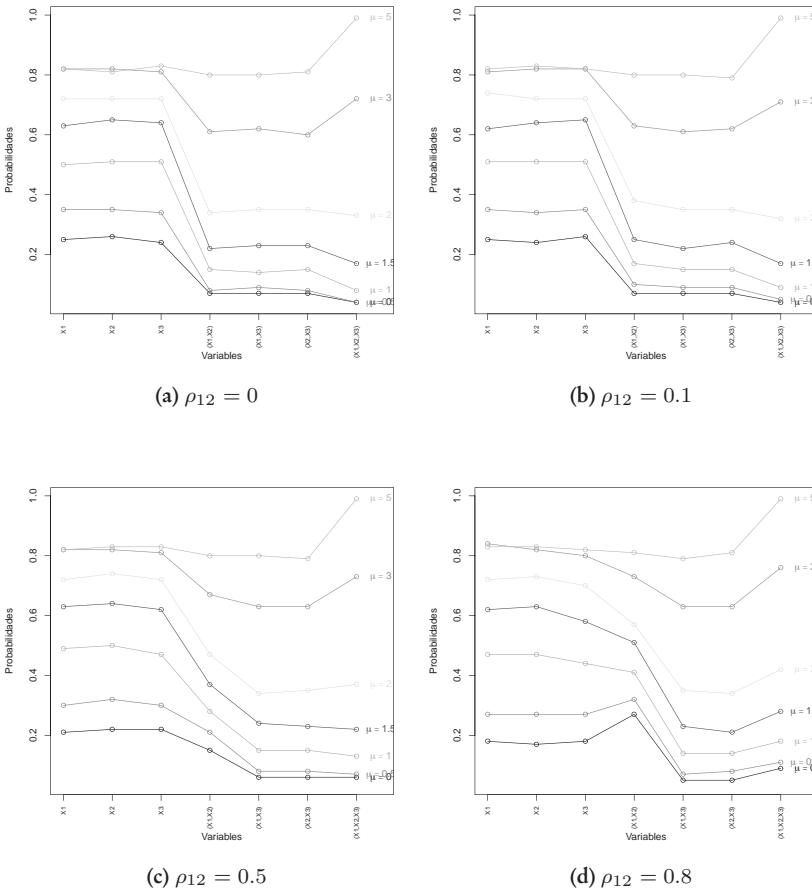


Figura 2: Perfiles de las probabilidades efectivas de detección de las variables según corrimientos en el vector de medias y correlación

6. CONCLUSIONES

Al analizar las tablas se observan algunos comportamientos en los datos, que se ven reflejados a través de las probabilidades realizadas en la metodología de Murphy (1987). Estos tipos de comportamientos se analizarán para determinar las conclusiones:

1. Como se observó la metodología de Murphy (1987) es una buena alternativa para la detección de las causas de una señal fuera de control en un proceso multivariado.
2. La efectividad de detección de la variable que ocasiono la señal fuera de control es mayor cuando se ha realizado un cambio grande en el vector de medias. Al realizar cambios muy pequeños la probabilidad es baja.
3. Para correlaciones altas entre las variables (X_1, X_2) , la probabilidad de detección es alta.
4. La detención efectiva en las variables de manera independiente, cambia cuando el corrimiento en media es pequeño o no hay y la correlación es alta.

Para estudios posteriores se podría aumentar el numero de variables, y tamaños de subgrupos mayores que uno, para aplicar la metodología y determinar su efectividad ante estos cambios. El estudio de la metodología de Murphy (1987) se centra en utilizar análisis discriminante de Fisher, el cual compara las medias de dos poblaciones multivariadas distribuidas normal p -variadas con la misma matriz de covarianzas. Sin embargo, al utilizar nuestra metodología de estudio consideramos correlaciones entre variables, e.d., cambios en la matriz de covarianzas sin modificar la regla del discriminante de Fisher. Un posible estudio sería evaluar qué pasa si se cambia la regla del discriminante a matrices de covarianzas distintas Díaz (2002). Das y Prakash (2008) comparan las propuestas realizadas Mason y cols. (1995), Murphy (1987), Hawkins (1993) y Doganaksoy y cols. (1991) utilizando pruebas de potencia sin tener en cuenta la fase en la que se encuentra cada proceso. Se pueden observar metodologías diferentes utilizando pruebas empíricas como las realizadas en este trabajo, teniendo en cuenta la fase en la que se encuentra el proceso.

BIBLIOGRAFÍA

- Alt, F. B. (1985). Multivariate quality control. En S. Kotz, N. L. Johnson, y C. R. Read (Eds.), *The encyclopedia of statistical sciences* (pp. 110–122). New York: John Wiley & Sons.
- Bersimis, S., Psarakis, S., y Panaretos, J. (2007). Multivariate statistical process control charts: An overview. *Quality and Reliability Engineering International*, 23(5), 517–543.
- Das, N., y Prakash, V. (2008). Interpreting the out-of-control signal in multivariate control chart - a comparative study. *The International Journal of Advanced Manufacturing Technology*, 37(9), 966–979.
- Doganaksoy, N., Faltin, F. W., y Tucker, W. T. (1991). Identification of out of control multivariate characteristic in a multivariable manufacturing environment. *Communications in Statistics - Theory and Methods*, 20(9), 2775–2790.
- Díaz, L. G. (2002). *Estadística Multivariada: Inferencia y Métodos* (Primera ed.). Bogotá: Universidad Nacional de Colombia, Departamento de Matemáticas y Estadística.
- Hawkins, D. M. (1991). Multivariate quality control based on regression adjusted variables. *Technometrics*, 33(1), 61–75.
- Hawkins, D. M. (1993). Regression adjustment for variables in multivariate quality control. *Journal of Quality Control*, 25(3), 170–182.
- Hotelling, H. (1947). Multivariate quality control - illustrated by the air testing of sample bombsights. En C. Eisenhart, M. W. Hastay, y W. A. Wallis (Eds.), *Techniques of statistical analysis* (pp. 111–184). New York: Mac-Graw Hill.
- Jackson, J. E. (1991). *A user guide to principal components*. New York: John Wiley & Sons.
- Maravelakis, P. E. (2003). *An investigation of some characteristics of univariate and multivariate control charts*. Tesis Doctoral, Department of Statistics,

Athenas University of Economics and Business.

- Mason, R. L., Tracy, N. D., y Young, J. C. (1995). Decomposition of T^2 for multivariate control chart interpretation. *Journal of Quality Technology*, 27(2), 99–108.
- Mason, R. L., Tracy, N. D., y Young, J. C. (1997). A practical approach for interpreting multivariate T^2 control chart signals. *Journal of Quality Technology*, 29(4), 396–406.
- Montgomery, D. C. (2004). *Control Estadístico de la Calidad* (Tercera ed.). México: Limusa Wiley.
- Murphy, B. J. (1987). Out of control variables with the T^2 multivariate quality control procedure. *Royal Statistical Society*, 36, 571–581.
- Roy, J. (1958). Step down procedure in multivariate analysis. *Annals of Mathematical Statistics*, 29, 1177–1187.
- Runger, G. C., Alt, F. B., y Montgomery, D. (1996). Contributors to a multivariate statistical process control chart signal. *Communications in Statistics. Theory and Methods*, 25(10), 2203–2213.
- Seber, G. A. F. (1984). *Multivariate observations*. New York: John Wiley.
- Timm, N. H. (1996). Multivariate quality control using finite intersection tests. *Journal of Quality Control*, 28(2), 233–243.
- Vargas, J. A. (2006). *Introducción al Control Estadístico de Calidad*. Bogotá: Universidad Nacional de Colombia, Departamento de Matemáticas y Estadística.
- Woodall, W. H., y Montgomery, D. C. (1999). Research issues and ideas in statistical process control. *Journal of Quality Technology*, 31(4), 376–386.
- Zambrano, A. J., y Zambrano, L. A. (2008). Evaluando la efectividad de la descomposición MYT para la interpretación de señales fuera de control en la carta T^2 . *Revista Tumbaga*, 3, 141–157.