

# ANÁLISIS Y EVALUACIÓN DE PROCESOS POR LOTES APLICANDO LA ESTADÍSTICA T<sup>2</sup> DE HOTELLING

## ANALYSIS AND EVALUATION OF APPLYING BATCH HOTELLING T<sup>2</sup> STATISTICS

<sup>1</sup>Raphael Félix Valencia Chacón

### Resumen

Los procesos utilizados en la industria se pueden dividir en dos grandes grupos: los procesos continuos y los procesos por lotes (batch). Las técnicas utilizadas para monitorear y controlar los procesos continuos no son directamente aplicables a los procesos por lotes, ya que estos últimos presentan una dimensión adicional con la que se tiene que tratar: el tamaño del lote. En la búsqueda de controlar la variabilidad de un proceso por lotes, se logró descomponer la variabilidad total de las observaciones de los lotes en dos fuentes independientes: una fuente que explique la variabilidad entre las mediciones realizadas en diferentes periodos de tiempo, es decir, entre lotes y otra fuente que explique la variabilidad entre las mediciones realizadas para un solo periodo de tiempo, es decir, dentro de lotes. Asimismo, al trabajar con procesos por lotes es posible que cada uno de los lotes tenga distinto vector de medias, alterando de esta manera la puesta en control del proceso; por ello, es necesario conocer el tipo de distribución normal que presentan los datos, y se buscará saber si los datos provienen de una distribución normal multivariada con un vector de medias iguales y con matriz de varianzas covarianzas común (Categoría 1) o provienen de una distribución normal multivariada con vector de medias diferentes y con matriz de varianzas covarianzas común (Categoría 2). Para implementar esta metodología, fue necesario usar la estadística T<sup>2</sup> modificada de Hotelling y para simplificar los cálculos de la estadística, se tuvo que elaborar un programa en el software libre R. De esta forma, se pudo llegar a un análisis adecuado de la variabilidad de un proceso por lotes. La aplicación de la metodología propuesta se efectuó en la empresa Frutos del Mar especializada en la exportación de moluscos bivalvos.

**Palabras clave:** Estadística T de Hotelling, procesos por lotes.

### Abstract

Manufacturing processes can be classified in two groups: continuous and batch production processes. The statistical techniques used to control continuous production processes are not applicable to batch production processes because the latter present an additional dimension to deal with: the batch size. Trying to control the variability of a batch production process, the total variability of the observations was divided in two independent sources: one to explain the variability between measures at different times, that is, between batches, and the other to explain the variability between measures at the same time, that is, within batches. Likewise, when working with batch production processes it is possible that each batch had a different mean vector. This situation affects the control setting of the process, so it is necessary to know if the data have a multivariate normal distribution with common mean vector and variance-covariance matrix (category 1) or if they have a multivariate normal distribution with different mean vectors and common variance covariance matrix (category 2). To implement this methodology was necessary to use the modified Hotelling T<sup>2</sup> statistics and to simplify the calculation of statistics had to develop a free software program R. This way it could reach a proper analysis of the variability of a batch process. The implementation of the proposed methodology was carried out in the company "Seafood" specialized in the export of bivalve molluscs.

**Keys words:** Hotelling T Statistic, process batch.

---

### 1. Introducción

El control estadístico de procesos es útil para lograr la estabilidad y mejorar la capacidad de procesos mediante la reducción de la variabilidad, la importancia del control de la variabilidad del proceso es que al estabilizarse la variable en estudio puede asumir una distribución

de probabilidad la cual puede ser predecible y, de esta manera, se pueden hacer predicciones estadísticas del proceso.

Los procesos por lotes son aquellos en los que la salida conduce a la producción de cantidades finitas de material. Para ello, se someten cantidades predefinidas de materia

<sup>1</sup> Médico Veterinario. Práctica Privada. EPG – UNALM. E-mail: [aphael@lamolina.edu.pe](mailto:aphael@lamolina.edu.pe)

prima a un conjunto ordenado de actividades de proceso sobre un periodo finito de tiempo y utilizan uno o más equipos.

En la literatura revisada sobre control de calidad, se encontró que se ha prestado menos atención a este tipo de procesos (por lotes). Las técnicas utilizadas para monitorear y controlar los procesos continuos no son directamente aplicables a los procesos por lotes, ya que estos últimos presentan un problema adicional con el que se tiene que tratar: el tamaño del lote.

La aplicación de los procesos por lotes no ha ido de la mano con el desarrollo de una teoría de control estadístico de procesos adecuada, que se ha enfocado mayormente a procesos continuos. Esto se debe a que el análisis estadístico de un proceso por lotes es muy especializado y las técnicas estadísticas no son muy difundidas, debido a un soporte matemático más complejo.

En un proceso por lote, en el que además de evaluar la variabilidad entre los lotes, se evalúan simultáneamente varias variables de calidad. La identificación de la existencia de *outliers* es más crítica, debido a que en la actualidad no se cuenta con herramientas estadísticas adecuadas para la identificación de la existencia de estos *outliers* y, por esta razón, es difícil establecer si el proceso por lotes se encuentra bajo control estadístico multivariado.

Al trabajar con procesos por lotes es posible que cada uno de los lotes tenga distinto vector de medias, alterando de esta manera la puesta en control del proceso; entonces es necesario conocer el tipo de distribución normal de los datos; ya que de esta manera se puede llegar a un análisis adecuado de la variabilidad de un proceso por lotes.

Actualmente, se tienen paquetes estadísticos (Minitab, SPSS) que permiten la construcción de gráficas de control multivariadas para procesos continuos; sin embargo, dichos procedimientos no pueden ser aplicados para procesos por lotes, debido a que la variabilidad del tamaño del lote es difícil de controlar con las metodologías conocidas. Por esta razón, se usará el programa R que es un programa de uso libre (libre distribución), mediante el cual se desarrollará una función que permita calcular un estadístico de prueba, el cual permitirá identificar el tipo de distribución normal multivariada que presenta un conjunto de datos (Categoría 1 o Categoría 2) y la identificación de la presencia de *outliers* (valores atípicos). Asimismo, se desarrollará una función que permita la descomposición de la variabilidad total del proceso.

En el presente trabajo de investigación, se realizó el análisis y evaluación de procesos por lotes aplicando la estadística  $T^2$  de Hotelling y modificándola, según las características de un proceso por lotes, cuyos objetivos son los siguientes:

- Identificar si un conjunto de datos provienen de una distribución normal multivariada con un vector de promedios común y una matriz de varianza-covarianza común (procesos por lotes de Categoría 1) o provienen de una serie de distribuciones normales multivariadas con diferente vector de promedios y una matriz de varianza-covarianza común (procesos por lotes de Categoría 2).
- Descomponer la variabilidad total de las observaciones de los lotes en dos fuentes independientes: una fuente que explique la variabilidad entre las mediciones realizadas en diferentes periodos de tiempo, es decir entre lotes y otra fuente que explique la variabilidad entre las mediciones realizadas para un solo periodo de tiempo, es decir, dentro de lotes.
- Obtener un estadístico de prueba  $T^2$  de Hotelling modificado, el cual al utilizar la variabilidad dentro de los lotes, permita verificar la existencia de *outliers* (valores atípicos) en la gráfica de control multivariada.
- Obtener un estadístico de prueba  $T^2$  de Hotelling modificado, el cual al utilizar la variabilidad entre los lotes permita verificar la existencia de *outliers* (valores atípicos) en la gráfica de control multivariada.

## 2. Revisión de literatura

El gráfico de control  $T^2$  de Hotelling es utilizado habitualmente para controlar simultáneamente varias características correlacionadas que indican la calidad de un único proceso productivo. Sin embargo, este gráfico es poco potente a la hora de detectar cambios pequeños o moderados en el proceso. Se describió el uso de señales adicionales de falta de control, en el cual se mostró las mejoras logradas en la potencia de la gráfica  $T^2$  de Hotelling y la frecuencia de aparición de dichas señales. (Aparisi y García, 2001).

Los gráficos de control de calidad más desarrollados hasta el momento para el control estadístico multivariado son los siguientes gráficos:  $T^2$  de Hotelling, Mewma, Mcusum. Estos gráficos son utilizados para el control estadístico de procesos, en los cuales es necesario controlar varias variables de manera simultánea. Para cada uno de los tipos de gráficos presentados, se analizaron sus características más importantes, propiedades, bondades y desventajas. (Barbiero, Flury, Pagura, Quaglino, Ruggieri, 2003).

Para la correcta interpretación de un valor  $T^2$  encontrado fuera de los límites de control, es necesario que este valor sea dividido en componentes, los cuales son independientes entre sí. Las gráficas de control multivariado, usando el estadístico  $T^2$  de Hotelling, son populares y fáciles de usar; sin embargo, la interpretación de los puntos fuera de control puede ser un problema. En este trabajo de investigación, se muestra que la interpretación de una señal proveniente del estadístico  $T^2$  de Hotelling es de gran ayuda si el valor correspondiente

es partido en dos partes independientes. La información acerca de qué característica es significativamente contribuyente para el análisis es rápidamente disponible a partir de esta descomposición. (Mason, Tracy y Young, 1995).

Mediante el uso de cartas de control, se pueden evaluar los parámetros de un proceso de producción y a partir de esa información determinar la capacidad del proceso. Considera que la meta final del Control Estadístico de Procesos (CEP) es eliminar la variabilidad en el proceso o reducirla al mínimo posible. Considera que el proceso se encuentra bajo control si la totalidad de los puntos muestrales se hallan dentro de los límites de control y se comportan de manera aleatoria. Si un punto se encuentra fuera de los límites se indica que el proceso se encuentra fuera de control. (Montgomery, Alt F. B. y Runger, 1996).

Para encontrar los límites de control en una carta T<sup>2</sup> de Hotelling en un proceso en el que se toman muestras de una sola observación, se desarrolla una metodología. Cuando se construyen gráficas de control para variables que están correlacionadas entre sí, es probable que al construir la carta de control para una variable no se encuentren señales de que el proceso está fuera de control; sin embargo, cuando se construyen cartas de control multivariado es probable que esta correlación se aprecie con la existencia de señales que indicarían que el proceso se encuentra fuera de control.

En una gráfica de una variable, la región de control tiene forma rectangular; mientras que, en el análisis multivariado, la región de control es representada por una elipse, en el que se evidencia una mayor restricción a las características para calificar si un proceso se encuentra fuera o bajo control. (Tracy, Young y Mason, 1992).

### 3. Materiales y métodos

#### 3.1 Clasificación de los procesos por lotes

Hay dos tipos básicos de procesos por lotes: procesos por lotes de Categoría 1 y procesos por lotes de Categoría 2.

**Categoría 1.-** En esta categoría, se asume que las observaciones de diferentes lotes provienen de una distribución normal multivariada común (con “p” variables), la cual es definida de la siguiente manera:

$$X \sim N_p(\mu, \Sigma)$$

Donde:  $\mu$  es un vector de media común y  $\Sigma$  es una matriz de varianza - covarianza común.

La existencia de pequeñas variaciones entre lotes puede ser aceptable; sin embargo, cuando esta variación es grande, los lotes atípicos son rechazados y se tiene que analizar en el proceso la causa de esa variabilidad.

**Categoría 2.-** Las observaciones de lotes de este tipo de procesos provienen de una serie de distribuciones normales multivariadas (con “p” variables), las cuales toman la misma matriz de varianzas – covarianzas, pero

posiblemente diferentes vectores de promedios, es decir, la distribución multivariada es definida de la siguiente manera:

$$X \sim N_p(\mu_i, \Sigma)$$

Donde:  $\mu_i$  para todo  $i=1, 2, \dots, k$ , es un vector media del  $i$ -ésimo lote.;  $k$  es el número de lotes,  $\Sigma$  es una matriz de varianza - covarianza común.

#### 3.2 Identificación del tipo de distribución normal multivariada de los datos

Para poder saber si los datos provienen de una distribución normal multivariada con el mismo vector de medias (Categoría 1) o provienen de una distribución normal multivariada con diferente vector de medias (Categoría 2), será necesario sustraer los correspondientes vectores media de lote estimado de cada uno de los conjuntos de datos de los lotes antes del análisis de los datos.

Con esta traslación, la diferencias de las medias entre los lotes son removidos y todos los datos son centrados en el origen. Trasladando los datos de lote también proveerá la estimación apropiada de la matriz de covarianza común.

Si usamos una media trasladada, todas las observaciones transformadas del proceso por lotes tendrían una distribución normal multivariada con la siguiente distribución:

$$Y \sim N_p(0, \Sigma)$$

Donde se tendrá un vector media cero y una matriz de covarianza  $\hat{\Sigma}$  común.

Los pasos a seguir para realizar el procedimiento de prueba serían los siguientes:

Paso 1: Centrar todos los datos de los lotes individuales. Esto se consigue restando cada valor de cada lote individual de su respectiva media de lote, tal como se muestra a continuación:

$$Y_{ij} = (X_{ij} - \bar{X}_i)$$

Donde:  $i=1, 2, \dots, k$  representa a los lotes,  $j=1, 2, \dots, n_i$  son las observaciones.

Con esta traslación, la variable  $Y_{ij}$  tiene la siguiente distribución:

$$Y \sim N_p(0, \Sigma)$$

Donde:  $\mu$  es un vector promedio igual a cero, con lo que cualquier posible diferencia de promedios ha sido removida debido a la traslación con el promedio cero y  $\Sigma$  es una matriz de varianza - covarianza común.

Paso 2: Usando los datos originales trasladados, calcular

el estimador de la matriz de varianza - covarianza (Sw) de la variación dentro de lotes, el cual está dado por:

$$S_w = \frac{1}{(N-k)} \sum_{i=1}^k \sum_{j=1}^m Y_{ij} Y_{ij}^T = \frac{1}{(N-k)} \sum_{i=1}^k \sum_{j=1}^m (X_{ij} - \bar{X}_i)(X_{ij} - \bar{X}_i)^T = \frac{SS_w}{N-k}$$

Nota: Para el cálculo de Sw, se construiría una función con el programa R que es un *software* de libre distribución.

Paso 3: Usar el estadístico de prueba T<sup>2</sup> y la distribución dada por:

Donde:  $\bar{X}_i$  : es el promedio para el i-ésimo lote,  $S_w$ : es el estimador de la matriz de varianza - covarianza de la variación dentro de lotes,  $\bar{X}$  : es el promedio general de todas las observaciones.

Paso 4: Usar la distribución dada por:

$$T^2 \approx \left[ \frac{p(k-1)(n-1)}{n[k(n-1)-p+1]} \right] F_{(p, nk-k-p+1)}$$

Donde:

$F_{[p, nk-k-p+1]}$ : representa la distribución F con parámetros p y (nk - k - p + 1).

Para un nivel de significación “α” dado, el Límite Superior de Control (LSC) para el estadístico T<sup>2</sup> es computado como:

$$LSC = \left[ \frac{p(k-1)(n-1)}{n[k(n-1)-p+1]} \right] F_{(\alpha, p, nk-k-p+1)}$$

Donde:  $F_{[\alpha, p, nk-k-p+1]}$ : es el α-ésimo cuartil superior de  $F_{[p, nk-k-p+1]}$ .

Paso 5: Plotear cada uno de los valores T<sup>2</sup> con el LCS establecido, para construir la gráfica de control.

Para realizar la prueba de hipótesis para la igualdad de medias de lotes, se asumirá que una observación del i-ésimo lote es distribuido como  $X \sim N_p(\delta_i, \Sigma)$ , y que una observación del j-ésimo lote es distribuido como  $X \sim N_p(\delta_j, \Sigma)$ .

En la forma de hipótesis estadística, se asume que la igualdad de las medias de lote es representado como:

$$H_0 : \delta_1 = \delta_2 = \dots = \delta_k$$

La hipótesis alternante es representada por:

$$H_1 : \delta_i \neq \delta_j \text{ para algun } i \neq j$$

Bajo H<sub>0</sub> cierta, se concluye que cada vector de observación X de cada lote proviene de una distribución normal multivariada con media común, es decir, lo siguiente:  $X \sim N_p(\delta, \Sigma)$ , cuyo proceso por lotes pertenece a la Categoría 1.

Bajo Ha cierta, se concluye que al menos un vector de observación X de cada lote proviene de una diferente distribución normal multivariada, es decir, lo siguiente:  $X \sim N_p(\delta_i, \Sigma)$ , cuyo proceso por lotes está definido por la Categoría 2.

Si en la gráfica de control no se encuentran puntos que se ubican fuera de los límites de control, se puede concluir con un nivel de significación establecido, es decir, los datos provienen de una distribución normal multivariada con media común y varianza común.

Si en la gráfica de control se encuentran puntos que se ubican fuera de los límites de control, se puede concluir con un nivel de significación establecido, esto es, los datos provienen de una distribución normal multivariada con diferente vector de medias y varianza común.

### 3.3 Clasificación de los procesos según el tamaño del lote

Según el tamaño del lote, los procesos por lotes se pueden clasificar:

#### Tamaño del lote = 1

Supongamos que recolectamos datos de “k” lotes tomados de cualquiera de los procesos por lotes: categoría 1 o categoría 2. Asumimos que el tamaño del lote es uno (es decir:  $n_i = 1, i=1, \dots, k$ ), de manera tal que una sola observación está hecha de “p” variables de cada lote.

De esta manera, el tamaño de la muestra total es  $n = \sum_{i=1}^k n_i = k$ . En este escenario, los *outliers* individuales pueden ser detectados computando el valor de T<sup>2</sup> para cada observación X, y comparando el valor con un LSC.

El correspondiente estadístico T<sup>2</sup> está dado como:

$$T^2 = (X - \bar{X})^T \Sigma^{-1} (X - \bar{X}) \quad (1)$$

Donde: “N” es el número total de observaciones, “p” es el número de variables, el vector  $\bar{X}$  y la matriz  $\Sigma$  son los estimadores comunes, respectivamente, de la media poblacional y de la matriz de varianza - covariancia en conjunto; y se representan de la siguiente manera:

$$\bar{X} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \vdots \\ \bar{X}_{(p+1)} \end{bmatrix}$$

El vector  $\bar{X}$  es de orden “p” x 1.

$$\Sigma = \begin{bmatrix} S_{i1}^2 & S_{i2} & S_{i3} & \dots & S_{ip} \\ S_{21} & S_{22}^2 & S_{23} & \dots & S_{2p} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ S_{p1} & S_{p2} & S_{p3} & \dots & S_{p2}^2 \end{bmatrix}_{(p \times p)}$$

La matriz de varianzas - covarianzas  $\hat{\Sigma}$  es de orden p x p, es decir, es cuadrática.

La matriz de varianzas - covarianzas de “n” observaciones y “p” variables puede ser obtenida de los datos originales de la siguiente manera:

$$\hat{\Sigma} = \frac{1}{n-1} \left( X - \frac{1}{n} X U \right) \left( X - \frac{1}{n} X U \right)^T$$

Donde: “n” es el número de observaciones; U es una matriz de “unos” y “X” es la matriz de datos de “n” observaciones y “p” variables, los cuales son expresados de la siguiente manera:

$$X = \begin{bmatrix} X_{11} & X_{12} & X_{13} & \dots & X_{1n} \\ X_{21} & X_{22} & X_{23} & \dots & X_{2n} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ X_{p1} & X_{p2} & X_{p3} & \dots & X_{pn} \end{bmatrix}_{(p \times n)}$$

La matriz de datos “X” es de orden p x n.

$$U = \begin{bmatrix} 1_{11} & 1_{12} & 1_{13} & \dots & 1_{1n} \\ 1_{21} & 1_{22} & 1_{23} & \dots & 1_{2n} \\ \cdot & \cdot & \cdot & \dots & \cdot \\ \cdot & \cdot & \cdot & \dots & \cdot \\ 1_{n1} & 1_{n2} & 1_{n3} & \dots & 1_{nn} \end{bmatrix}_{(n \times n)}$$

La matriz de “unos” es de orden n x n, es decir, es cuadrática.

El Límite Superior de Confianza (LSC) es dado como:

$$LSC = \left[ \frac{(n-1)^2}{n} \right] B_{[\alpha, p/2, (n-p-1)/2]} \quad (2)$$

Donde:  $B_{(p/2, (n-p-1)/2)}$ : representa la distribución beta con parámetros  $(p/2)$  y  $((n-p-1)/2)$ .

“a”: es el nivel de significación establecido por el investigador.

### 3.4 Tamaño del lote mayor que 1

En los procesos de lotes categoría 1, se analiza la existencia de *outliers* individuales, en el cual se usa el estadístico T<sup>2</sup> en la ecuación 1 y el correspondiente LSC en la ecuación 2. Sin embargo, el tamaño de muestra N ya no se reduce a k pero tiene la forma general:

$$n = \sum_{i=1}^k n_i$$

De las observaciones que podrían ser tomadas de diferentes distribuciones por los lotes en el proceso de lotes categoría 2, la definición y detección de los *outliers* cambia de aquellos dados por la categoría 1.

Con los lotes, la media muestral en su conjunto en la ecuación 1 es equivalente al promedio ponderado de las medias de lote, definida como:

$$\bar{X} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}}{n} = \frac{\sum_{i=1}^k n_i \bar{X}_i}{n} \quad (3)$$

Donde:  $X_{ij}$ : es la j-ésima observación en i-ésimo lote,  $\bar{X}_i$ : representa el vector media de i-ésimo lote.

Así mismo, el estimador de la matriz de covarianza usado en la ecuación 1 mide no solo la variación del proceso dentro del lote, sino también la variación del proceso debido a las diferencias entre los lotes.

El estimador  $\hat{\Sigma}$  es la variación total;  $M_T$  dividido por (n - 1) y es escrito como:

$$\hat{\Sigma} = \frac{1}{n-1} \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}) (X_{ij} - \bar{X})^T = \frac{M_T}{n-1} \quad (4)$$

La variación total también puede ser escrita como:

$$M_T = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i) (X_{ij} - \bar{X}_i)^T + \sum_{i=1}^k n_i (\bar{X}_i - \bar{X}) (\bar{X}_i - \bar{X})^T \quad (5)$$

$M_T$  = Variación dentro del lote + Variación entre lotes

$$M_W = \text{Variación dentro del lote} = \sum_{i=1}^k \sum_{j=1}^{n_i} (X_{ij} - \bar{X}_i) (X_{ij} - \bar{X}_i)^T$$

$$M_B = \text{Variación entre lotes} = \sum_{i=1}^k n_i (\bar{X}_i - \bar{X}) (\bar{X}_i - \bar{X})^T$$

En el proceso por lotes de categoría 1, se asume que la variación entre lotes ( $M_B$ ) es mínima y es estrictamente debido a las fluctuaciones al azar. Así, los elementos correspondientes de  $M_T$  y  $M_W$  son muy similares en tamaño, es decir, se puede afirmar que en el proceso de lote categoría 1 se tiene:  $M_T = M_W$ .

Sin embargo, en el proceso por lotes categoría 2, la variación entre lotes ( $SS_B$ ) puede ser larga y producir una diferencia considerable entre los correspondientes elementos  $SS_T$  y  $SS_W$ , es decir, se puede afirmar que en el proceso de lote categoría 2 se tiene:  $M_T = M_W + M_B$ .

Los lotes individuales también podrían contener *outliers* estadísticos, que tienden a influenciar grandemente la estimación de la matriz de covarianza para tamaños de muestra pequeñas.

Si se descompone el valor de la matriz  $\hat{\alpha}$ , se llega a las siguientes expresiones:

$$\text{Como: } \Sigma = \frac{M_T}{n-1}, \text{ entonces: } \Sigma_W = \frac{M_W}{n-k} \text{ y } \Sigma_B = \frac{M_B}{k-1}$$

Donde: La matriz varianza-covarianza para  $\hat{\alpha}_W$  y  $\hat{\alpha}_B$ , pueden ser obtenidas respectivamente de la siguiente manera:

$$\Sigma_W = \frac{1}{n-k} \left( x - \frac{1}{m} x U^* \right) \left( x - \frac{1}{m} x U^* \right)^T$$

$$\Sigma_B = \frac{1}{k-1} \left( \frac{1}{m} x U^* - \frac{1}{n} x U \right) \left( \frac{1}{m} x U^* - \frac{1}{n} x U \right)^T$$

Donde: “n” es el número total de observaciones, “k” el número de lotes, “m” es el número de sub muestras (todas del mismo tamaño) obtenidas de cada lote “k”,  $U^*$  es una matriz (n x n) de unos (1), construida de la siguiente manera:

$$U^* = \begin{bmatrix} \begin{bmatrix} 1 & 1 & \dots & 1_{(m)} \end{bmatrix} & & & \\ \dots & & & \\ \begin{bmatrix} 1 & 1 & \dots & 1_{(m)} \end{bmatrix} & 0 & \dots & 0 \\ \dots & & & \\ 0 & \begin{bmatrix} 1 & 1 & \dots & 1_{(m)} \end{bmatrix} & \dots & 0 \\ \dots & \begin{bmatrix} 1 & 1 & \dots & 1_{(m)} \end{bmatrix} & & \\ \dots & & & \\ 0 & \dots & & \begin{bmatrix} 1 & 1 & \dots & 1_{(m)} \end{bmatrix} \\ \dots & & & \begin{bmatrix} 1 & 1 & \dots & 1_{(m)} \end{bmatrix} \end{bmatrix}_{(nm)}$$

El cálculo de las matrices, las cuales se requieren para estimar tanto la variabilidad dentro de los lotes ( $\hat{\alpha}_W$ ) como la variabilidad entre los lotes ( $\hat{\alpha}_B$ ), respectivamente, es un proceso bastante complejo. Por ello, se desarrolló un programa en el *software* libre R, que nos permitió simplificar el proceso de cálculo y con el cual se pueden estimar rápidamente tanto  $\hat{\alpha}_W$  como  $\hat{\alpha}_B$ , además de lograr descomponer así la variabilidad total de las observaciones.

Si esta observación de los lotes no es tomada en cuenta, luego al construir la matriz de varianzas - covariancias

( $\hat{\alpha}$ ) para el total de observaciones, se podrá llegar a una estimación errónea de la variabilidad total, debido a que el cálculo de matriz de varianza - covarianza está basado en las desviaciones tomadas desde la media en su conjunto.

### 3.5 Gráfica de control multivariada en procesos por lotes de Categoría 1

Para un proceso por lotes de categoría 1, se asume que las observaciones de diferentes lotes provienen de una distribución normal multivariada común (con “p” variables), la cual es definida de la siguiente manera:

$$X \sim N_p (\mu, \Sigma)$$

Donde:  $\mu$  es un vector de media común y  $\Sigma$  es una matriz de varianza - covarianza común.

Para la construcción de la gráfica de control multivariada para un proceso por lotes de Categoría 1, se tendrá que el estadístico de prueba esta dado por:

$$T^2 = \left( \overline{X}_i - \overline{X} \right)^T \Sigma_W^{-1} \left( \overline{X}_i - \overline{X} \right) \quad (6)$$

Donde:  $\overline{X}_i$ : es el promedio para el i-ésimo lote,  $\hat{\alpha}_W = M_W / (n - k)$  es la estimación de la matriz de varianza-covarianza en (5),  $\overline{X}$  es el promedio general con la ecuación (3).

Para un nivel de significación “ $\alpha$ ” dado, el LSC para el estadístico  $T^2$  es definido como:

$$LSC = \left[ \frac{p(k-1)(n-1)}{n[k(n-1) - p + 1]} \right] F_{(\alpha, p, nk - k - p + 1)} \quad (7)$$

Donde:  $F_{[p, k - k - p + 1]}$  representa la distribución  $F$  con parámetros  $p$  y  $(nk - k - p + 1)$ .

El valor de  $T^2$  en la ecuación 6 es calculado para cada una de las  $k$  medias de lotes y comparadas con el LSC en la ecuación 7. Las medias de lote con los valores de  $T^2$  que exceden el LSC son declarados *outliers*. La existencia de estos *outliers* indica que el proceso no se encuentra bajo control estadístico y se tiene que proceder con la identificación de las causas asignables de variación para la mejora del proceso.

### 3.6 Gráfica de control multivariada en procesos por lotes de Categoría 2

Para un proceso por lotes de categoría 2, las observaciones de lotes de este tipo de procesos provienen de una serie de distribuciones normales multivariadas (con “p” variables) tomando la misma matriz de varianzas - covarianzas, pero posiblemente diferentes vectores de promedios, es decir la distribución multivariada es definida de la siguiente manera:

$$X \sim N_p (\mu_i, \Sigma)$$

Donde:  $\mu_i$ , para todo  $i = 1, 2, \dots, k$ , es un vector media del  $i$ -ésimo lote y  $\Sigma$  es una matriz de varianza - covarianza común.

Para la construcción de la gráfica de control multivariada para un proceso por lotes de categoría 2, se tendrá que el estadístico de prueba esta dado por:

$$T^2 = \left( \bar{X}_i - \bar{X} \right)^T \Sigma_B^{-1} \left( \bar{X}_i - \bar{X} \right) \quad (8)$$

Donde:  $\bar{X}_i$ : es el promedio para el  $i$ -ésimo lote,  $\hat{\Sigma}_B = \mathbf{M}_B / (k-1)$  es la estimación de la matriz de varianza-covarianza definido en la ecuación 5,  $\bar{X}$ : es el promedio general calculado con la ecuación 3.

Para un nivel de significación " $\alpha$ " dado, el LSC para el estadístico T<sup>2</sup> es calculado como:

$$LSC = \left[ \frac{(k-1)^2}{k} \right] B_{(\alpha, p/2, (k-p-1)/2)} \quad (9)$$

Donde:  $k$ : es el número de lotes,  $B_{(\alpha, p/2, (k-p-1)/2)}$ : representa a la distribución Beta que tendrá como parámetros  $(p/2)$  y  $(k-p-1)/2$ .

**4. Resultados**

Para la aplicación de la metodología, se recogió datos de la empresa Frutos del Mar, la cual se dedica a la exportación de productos hidrobiológicos. Uno de los productos que se exporta es el molusco bivalvo (Tellina); del cual es necesario llevar un control de tres variables críticas, las cuales son temperatura (en °C), oxígeno (en ppm) y salinidad (en ppm) para el producto Tellina (*Donax trunculus*). Con el programa elaborado en el software R, se obtienen las siguientes matrices:

**Tabla 1.** Matriz de Varianza – Covarianza del total de Observaciones ( $\hat{\Sigma}$ ).

Variables	Temperatura (°C)	Salinidad (ppm)	Oxígeno (ppm)
Temperatura (°C)	4.3333077	1.4479615	0.4885641
Salinidad (ppm)	1.4479615	2.0810192	0.3922564
Oxígeno (ppm)	0.4885641	0.3922564	0.2144615

Fuente: Elaboración propia en base a datos de campo, 2010.

**Tabla 2.** Matriz de Varianza – Covarianza dentro del lote (Sw).

Variables	Temperatura (°C)	Salinidad (ppm)	Oxígeno (ppm)
Temperatura (°C)	2.464333	0.5571667	0.1010833
Salinidad (ppm)	0.5571667	1.3490833	0.1055833
Oxígeno (ppm)	0.1010833	0.1055833	0.107666

Fuente: Elaboración propia en base a datos de campo, 2010.

**Tabla 3.** Matriz de Varianza – Covarianza entre lotes ( $\Sigma_B$ ).

Variables	Temperatura (°C)	Salinidad (ppm)	Oxígeno (ppm)
Temperatura (°C)	10.563222	4.417278	1.780167
Salinidad (ppm)	4.417278	4.520806	1.347833
Oxígeno (ppm)	1.780167	1.347833	0.5704444

Fuente: Elaboración propia en base a datos de campo, 2010.

Posteriormente, de acuerdo con la metodología propuesta, se tiene que trasladar las observaciones a un origen común. Esto se logra utilizando la siguiente expresión:

$$Y_{ij} = (X_{ij} - \bar{X}_i)$$

Se procedió a probar si los datos de la muestra recogida correspondían a un proceso por lotes de categoría 1 o a un proceso por lotes de categoría 2.

La hipótesis planteada fue la siguiente:

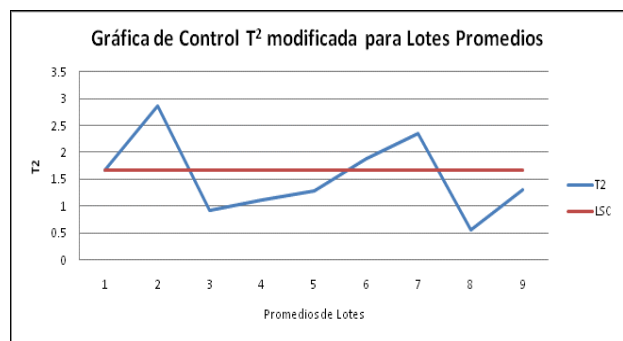
$$H_0 : \delta_1 = \delta_2 = \dots = \delta_k$$

Cada observación de lote viene de la misma distribución multivariada.

$$H_1 : \delta_i \neq \delta_j \text{ para algun } i \neq j$$

Al menos una observación de lote viene de diferentes distribuciones multivariadas.

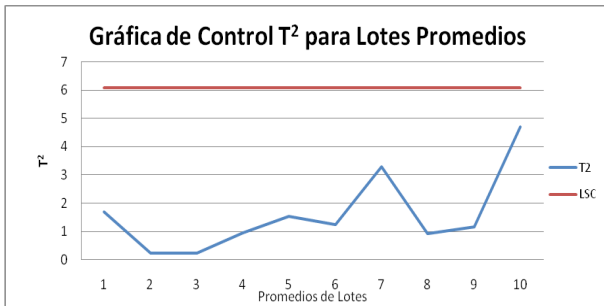
Para un nivel de significación del 10%, se construyó la siguiente Gráfica de Control de T<sup>2</sup> modificada de Hotelling:



**Figura 1.** Gráfica de Control T<sup>2</sup> modificada para lotes promedios.

Fuente: Elaboración propia, 2010.

Para un Límite Superior de Control (LSC) igual a 1.66825227 y un nivel de significación del 10%, se pudo afirmar que existen evidencias estadísticas para rechazar la hipótesis planteada (Ho) es decir se concluyó que al menos una observacion de lote vino de diferentes distribuciones multivariadas. Se procedió a construir la gráfica de control multivariado para un proceso por lotes de categoría 2.



**Figura 2.** Gráfica de Control  $T^2$  modificada para lotes promedios\_Tellina.

Fuente: Elaboración propia, 2010.

En la Figura 2, para un LSC igual a 6.09076267 calculado con la expresión (13) y un nivel de significación del 5%, se pueden observar la NO existencia de *outliers*, debido a que no se encontraron valores  $T^2$  que superaron al LSC; de esta manera se puede llegar a concluir que el proceso por lotes de categoría 2 si se encuentra bajo control.

## 5. Conclusiones

1. Se utilizó el estadístico  $T^2$  de Hotelling modificado, para descomponer la variabilidad total del proceso en dos fuentes independientes: una fuente que explique la variabilidad entre las mediciones realizadas en diferentes periodos de tiempo, es decir, entre lotes y otra fuente que explique la variabilidad entre las mediciones realizadas para un solo periodo de tiempo es decir dentro de lotes.
2. Se utilizó el estadístico  $T^2$  de Hotelling modificado, para verificar la existencia de valores extremos cuando se utiliza la variabilidad dentro de los lotes, encontrándose la presencia de valores que sobrepasan el límite de control superior.
3. Se utilizó el estadístico  $T^2$  de Hotelling modificado, para verificar la existencia de valores extremos cuando se utiliza la variabilidad entre los lotes, no encontrándose la presencia de valores que sobrepasan el límite de control superior.
4. Para poder determinar si una matriz de datos  $X$  que tiene una distribución normal multivariada presenta un vector de medias común (categoría 1) o presenta un vector de medias diferentes (categoría 2), será necesario realizar la traslación de los datos de los lotes a un origen común substrayendo los vectores medias individuales de cada lote de su correspondiente vector promedio de lote de observaciones. Con esta traslación, se produce un conjunto global de datos, los cuales tendrán un vector de medias igual a cero y una matriz de varianza-covarianza común.

## 5. Literatura citada

- Aparisi y Garcia (2001).** “Aumento de la potencia del gráfico de control multivariante  $T^2$  de Hotelling utilizando señales adicionales de falta de control”. Universidad Politécnica de Valencia (España)
- Barbiero, Flury, Pagura, Quaglino, Ruggieri (2003).** “Control estadístico de procesos multivariados mediante graficos de control multivariados  $T^2$  de Hotelling, Mewma y Mcusum”. Universidad Nacional de Rosario (Argentina)
- Chairman, Schmidt y Koelling (1997).** “A performance analysis of the minimax multivariate quality”
- Mason, R. L., Tracy N. D. y Young J. C. (1995).** “Descomposition of  $T^2$  for Multivariate Control Chart Interpretation”. Journal of Quality Technology 27, pp 99 - 108
- Montgomery, D. C., Alt, F. B. and Runger, G. C. (1996).** “Contributors to a Multivariate Statistical Process Control Chart Signal”, Comm. Statist. Theory Methods, 25, pp. 2203-2213.
- Nedumaran and Pignatiello (2004).** “On constructing  $T^2$  Control chart for retrospective examination”, University Florida State (USA).