



**Barbiero, Cristina**

**Flury, Maria Isabel**

**Pagura, José Alberto**

**Quaglino, Marta B.**

**Ruggieri, Marta A.**

*Instituto de Investigaciones Teóricas y Aplicadas. Escuela de Estadística.*

## **EFFECTO DE LA FALTA DE NORMALIDAD EN GRÁFICOS DE CONTROL PARA EL COEFICIENTE DE VARIACIÓN**

### **1- INTRODUCCIÓN**

Cuando se usa el término calidad, se suele imaginar un excelente producto o servicio, que cumple o rebasa las expectativas del usuario. Estas expectativas se basan en el uso que se le pretende dar y también en el precio de venta.

Para el consumidor la calidad es un factor importante de decisión y para las empresas, es fundamental para su crecimiento y posicionamiento en el mercado. Todas las acciones planeadas o sistemáticas necesarias para proporcionar una confianza adecuada de que el producto o servicio va a satisfacer determinados requisitos de calidad, se llama *aseguramiento de la calidad*. Implica asegurarse de que la calidad es la que debería ser.

Las empresas que cuentan con programas efectivos de aseguramiento de la calidad pueden lograr una mayor productividad y penetración en el mercado al mismo tiempo que pueden disminuir costos generales de fabricación y de servicios. El rápido avance de la tecnología genera permanentemente un gran número de productos y servicios nuevos, lo que hace que el usuario disponga de un mayor abanico de posibilidades. Alcanzar la calidad en el ambiente moderno de los negocios y de la manufactura no resulta tan sencillo, pero puede lograrse a través de un adecuado control.

Se denomina *control estadístico de calidad* al conjunto de técnicas y de métodos operativos que permiten tomar acciones correctivas apropiadas para desarrollar, mantener y mejorar la calidad de un producto o servicio.

El *control estadístico de calidad* se compone de un grupo de herramientas que incluyen el control estadístico de procesos (SPC), planes de muestreo, diseño de experimentos, análisis de confiabilidad y herramientas de diagnóstico, entre otros.

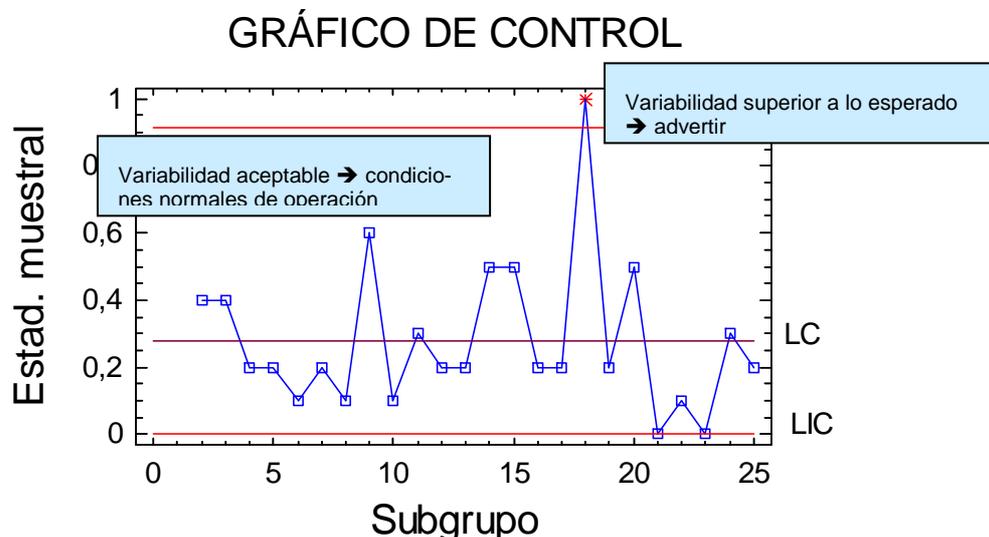
Tradicionalmente el control de calidad fue usado en procesos de fabricación, pero actual-



mente se ha extendido a otras áreas como educación y salud y a una variedad de aplicaciones sociales.

Una herramienta fundamental en el *SPC* son los gráficos de control, que constituyen un método de análisis y presentación de datos que permite evaluar cuándo las variaciones observadas en la calidad son mayores que las que causaría la casualidad. Son una representación gráfica en el tiempo de una estadística medida o calculada a partir de subgrupos racionales o muestras. El gráfico tiene una línea central (LC) que representa el valor medio o esperado de la estadística bajo condiciones normales de operación, que corresponden al denominado "estado de control estadístico" y otras dos líneas horizontales denominadas límite superior (LSC) e inferior de control (LIC) que abarcan un rango dentro del cual es altamente probable que se encuentre dicha estadística, si el proceso funciona bajo condiciones normales. El valor de probabilidad que debe abarcarse es fijado por el usuario, pero es típico establecerlo en niveles superiores al 99%. Mientras los puntos graficados se encuentren entre los límites, se considera que el proceso está bajo control, en un estado estable y predecible. En cambio si un punto cae fuera, se interpreta como una evidencia de que una causa asignable de variación está llevando al proceso fuera de control, excediéndose de la variación natural esperada (ver Gráfico 1). Los límites de control permiten juzgar la importancia de las variaciones en la calidad del producto o servicio y no deben confundirse con los de especificación, que son los límites admisibles de una característica de calidad para cada unidad individual de un producto, que depende únicamente de especificaciones técnicas o requerimientos del mercado y no de consideraciones probabilísticas.

**Gráfico 1.** Esquema general de un gráfico de control.





En particular si la variable de calidad que se define es de tipo cuantitativa, los gráficos son conocidos en la literatura como gráficos de control por variables y ellos permiten visualizar las variaciones que se presentan en el "*nivel medio*" y en "*la variabilidad*" de cierta característica de calidad. Los gráficos de control por variables más usados son los gráficos de promedios y rangos ( $\bar{X}$ ,  $R$ ) y los de promedios y desvíos ( $\bar{X}$ ,  $S$ ) que son útiles en muchos escenarios siempre y cuando el proceso bajo control se caracterice por tener promedio y variancia estables. Ésta es una propiedad imprescindible y sólo en estos casos puede utilizarse el gráfico  $\bar{X}$  para monitorear la media del proceso y el gráfico de rangos ( $R$ ) o de desvíos ( $S$ ) para monitorear la variabilidad.

Una situación en la que estos supuestos no se cumplen, puede darse cuando la variabilidad no se mantiene constante, porque ella cambia cuando la media cambia. En ese caso se puede llevar adelante un análisis en el cual se tenga en cuenta la dependencia entre variancia y promedio, calculando pivotes que midan la variabilidad inherente del proceso sin la interferencia de la media que varía. Se entiende por pivote una estadística con distribución conocida e independiente de los parámetros desconocidos. La naturaleza de la relación entre el promedio y la variancia del proceso, dará la pauta de cómo delinear una posible solución al problema de la variabilidad no constante. El vínculo entre ambos parámetros, suele expresarse en términos de su cociente, que es una medida de variabilidad relativa denominada: *coeficiente de variación poblacional* ( $\gamma$ ) que se define como el cociente entre  $\sigma$  y  $\mu$ . Este coeficiente resulta constante para el caso particular de una relación directamente proporcional entre la desviación estándar  $\sigma$  y la media  $\mu$ .

El coeficiente de variación, se estima a menudo por simple expansión, mediante el coeficiente de variación muestral  $W = S/\bar{X}$ . Ésta es una medida útil en muchas áreas: en finanzas y seguros para determinar riesgos relativos; en Bancos para controlar movimientos de cuentas mediante el manejo del costo de comisiones por transacción o tasas de interés por el dinero depositado; en el área médica y de diagnóstico para determinar precisión de mediciones; en el área industrial para comparar métodos de medición diferentes.

En los casos en los cuales la variabilidad es proporcional a la media, para monitorear el funcionamiento del proceso puede utilizarse un gráfico de control para el coeficiente de variación.



El Gráfico de Control para el Coeficiente de Variación representa una aplicación no tradicional de los Gráficos de Shewhart y su diseño presenta algunas dificultades particulares, por la complejidad de las propiedades distribucionales del coeficiente de variación  $W$ .

## 2 – OBJETIVO

En este trabajo se presenta un gráfico de control para el CV usando límites probabilísticos, bajo el supuesto de que las observaciones tienen distribución normal. También se evalúa la robustez de esta estrategia frente a alejamientos suaves del supuesto de normalidad.

## 3 – METODOLOGÍA

En el caso general de gráficos de control por variables en los cuales la dispersión no depende de la media, ambos gráficos ( $\bar{X}, R$  o  $\bar{X}, S$ ) deben interpretarse conjuntamente. El gráfico  $R$  (o  $S$ ) está afectado únicamente por cambios en la dispersión de los datos, mientras que el gráfico  $\bar{X}$  se ve afectado tanto por cambios en la media del proceso como por su dispersión. Se presenta una situación fuera de control cuando uno o más puntos se encuentran fuera de los límites, o bien cuando los puntos muestran algún patrón de comportamiento no aleatorio. Si ambos gráficos ( $\bar{X}, R$  o  $\bar{X}, S$ ) presentan un patrón no aleatorio, deben eliminarse primero las causas atribuibles en el gráfico  $R$  o  $S$ , ya que esto podría eliminar automáticamente el patrón en el gráfico  $\bar{X}$ . No debe interpretarse el gráfico de medias si el de variabilidad se encuentra fuera de control.

Cuando la variabilidad depende de la media los gráficos mencionados no son adecuados, ya que un punto fuera de límites podría no deberse a una causa asignable de variación sino precisamente a la dependencia mencionada. En el caso particular en el cual la desviación estándar es proporcional a la media, es adecuado el uso de un gráfico de control para el CV. Para su construcción se seleccionan subgrupos racionales, se calcula el promedio y la desviación estándar para cada subgrupo y se grafica el coeficiente de variación muestral  $W = S/\bar{X}$ .

Este gráfico es una secuencia sucesiva de los coeficientes de variación muestrales ordenados en el tiempo. Contiene tres líneas: una central, que representa el valor medio de la característica de calidad cuando el proceso está bajo control, y dos que constituyen los límites de control superior e inferior. Un tema importante a considerar es cómo seleccionar el mejor



estimador del CV poblacional para determinar la línea central del gráfico. Generalmente la línea central se corresponde con el valor esperado de la característica graficada, sin embargo en el gráfico para el coeficiente de variación todos los momentos del coeficiente de variación muestral  $W$  son infinitos y por lo tanto no puede determinarse ni la línea central ni los límites de control en la forma tradicional. Este problema se resuelve utilizando, no el valor esperado del coeficiente de variación muestral, sino el valor del coeficiente de variación poblacional ( $\gamma$ ) cuando el proceso está bajo control. Muchas veces, en la práctica, suele desconocerse este valor. Para estimarlo, se han propuesto tres posibles estadísticas:

- a- El promedio simple de los coeficientes de variación de cada muestra

$$\bar{W} = \sum_{j=1}^m W_j / m$$

- b- La raíz cuadrada de la media ponderada de los coeficientes de variación muestrales elevados al cuadrado

$$\hat{\gamma} = \sqrt{\frac{\sum_{i=1}^m (n_i - 1) W_i^2}{\sum_{i=1}^m (n_i - 1)}} \quad \text{con } n_i \text{ tamaño de } i\text{-ésima muestra.}$$

- c- El estimador máximo verosímil bajo el modelo normal.

La elección entre estos estimadores fue investigada por Connett and Lee (1990) y Kang et al (2007). Ellos recomiendan utilizar el  $\hat{\gamma}$  de la opción b-, pues presenta mayor eficiencia relativa respecto de  $\bar{W}$  y del estimador máximo verosímil.

Otro punto importante a considerar al momento de la implementación del gráfico, es que para la determinación de los límites de control no existen tablas disponibles para el usuario y las distribuciones de probabilidad de los estadísticos no son simples ni están muy divulgadas. Con respecto a la distribución de probabilidad de  $W$  se han dado distintas propuestas: aproximación de McKay (1932), basada en la distribución Chi-cuadrado; aproximación de Iglewicz (1967) y Verrill (2003), basada en la distribución  $t$  no centrada y aproximación de Kang (2007) basada en la distribución Canónica.

Para este trabajo, se calcularon límites probabilísticos bilaterales con las tres propuestas mencionadas, teniendo en cuenta distintos tamaños muestrales (5, 10 y 15) y distintos valores del coeficiente de variación poblacional (0,05; 0,10 y 0,15). Los límites se fijaron de modo tal que la probabilidad de excederlos (tasa de falsas alarmas) sea de 1/370 cuando el proceso está bajo control, valor similar al valor estándar utilizado en los gráficos de  $\bar{X}, R$  o



$\bar{X}, S$  con límites en  $3\sigma$ . Los resultados se muestran en la Tabla 1

Tabla 1. **LÍMITES DE CONTROL PARA EL GRÁFICO DE CONTROL DEL CV**

Tamaño muestra	CV	Aprox. Canónica		Aprox. Chi-cuadrada		t no centrada	
		LI	LS	LI	LS	LI	LS
n=5	$\gamma = 0.05$	0.00812	0.10587	0.00813	0.10580	0.00816	0.10630
	$\gamma = 0.10$	0.01621	0.21414	0.01620	0.21370	0.01629	0.21512
	$\gamma = 0.15$	0.02390	0.32850	0.02420	0.32600	0.02520	0.32760
n=10	$\gamma = 0.05$	0.01855	0.08696	0.01880	0.08690	0.01855	0.08696
	$\gamma = 0.10$	0.03701	0.17518	0.03740	0.1750	0.03701	0.17518
	$\gamma = 0.15$	0.0552	0.26910	0.05580	0.26540	0.05520	0.27010
n=15	$\gamma = 0.05$	0.02391	0.07949	0.04262	0.07945	0.02391	0.07949
	$\gamma = 0.10$	0.04770	0.15986	0.04770	0.15971	0.04770	0.15986
	$\gamma = 0.15$	0.07020	0.24090	0.07120	0.24164	0.07119	0.24165

Los límites obtenidos usando las tres distribuciones propuestas (canónica,  $\chi^2$  y  $t$  no centrada), son muy similares. Sólo presentan alguna diferencia en el orden del cuarto decimal, para los mayores valores de  $n$  y  $CV$  considerados ( $n=15$  y  $CV=0.15$ ). Sin embargo las dificultades de cálculo de los percentiles con cada distribución propuesta, no son las mismas.

Dados los resultados obtenidos, para evaluar la performance del gráfico para el CV, se consideró sólo la distribución canónica, utilizando como medida de referencia el valor de la *Longitud de corrida promedio (ARL)*.

El *ARL* es el número medio de puntos que deben graficarse desde que se produce una salida de control hasta que el gráfico la detecta. Son deseables valores altos de *ARL* cuando el proceso está bajo control y valores bajos cuando se sale de control. Para cada tamaño de subgrupo y valor del CV poblacional se obtiene una determinada curva *ARL*.

Se calcularon las curvas *ARL* bajo el supuesto de normalidad y bajo el de alejamientos suaves de la misma. A fin de comprobar el efecto de este alejamiento, se usaron los límites establecidos para la distribución normal y se estimaron los *ARL* a través de simulación de poblaciones normales contaminadas con un 20% de valores normales, con desvíos incrementados en un 5%, 10%, 15% y 20%. Como ilustración sólo se muestra el gráfico con los



ARL que corresponden a los distintos niveles de contaminación simulados, manteniendo fijos  $n=15$  y  $CV= 0.05$ .

**¡Error! No se pueden crear objetos modificando códigos de campo.**

Cuando el proceso se encuentra bajo control ( $CV_0 = 0.05$ ), alejamientos cada vez más pronunciados de la normalidad, se traducen en incrementos cada vez mayores en la tasa de falsas alarmas, llegando a producirse aumentos del 40 % en la tasa elegida para el diseño de la estrategia ( $1/370 = 0.0027$ ). Por el contrario, cuando el proceso se aparta de la situación bajo control ( $CV>0.05$ ), cuanto mayor es el alejamiento de la normalidad, más efectivo es el gráfico para detectar salidas de control, ya que los ARL son menores.

#### **4 – CONSIDERACIONES FINALES**

El gráfico de control para el CV es una herramienta útil para controlar procesos en los cuales los clásicos diagramas Shewhart no son aplicables, porque la variabilidad cambia cuando la media cambia.

El diseño de este gráfico no resulta sencillo debido a que no es factible el uso de los métodos convencionales para la determinación de los límites de control. Requiere el conocimiento de la distribución de probabilidad del coeficiente de variación muestral. Este tema ha sido abordado por diferentes autores cuyas propuestas se tomaron como base para obtener los límites probabilísticos del gráfico. Los resultados obtenidos fueron similares para las tres propuestas analizadas.

La performance del gráfico fue analizada a través de los ARL, usando la distribución Canónica. También se evaluó la robustez de la estrategia de control frente a alejamientos del supuesto de normalidad, hallando que el parámetro más afectado es la proporción de falsas alarmas.

#### **V - REFERENCIAS BIBLIOGRÁFICAS**

BARBIERO, C.; FLURY, M.; PAGURA, J.; QUAGLINO, M.; RUGGIERI, M. (2009). "Control charts for the coefficient of variation". 57<sup>th</sup> Session of the International Statistical Institute.



Durban, Sudáfrica.

CHANG W. KANG et all (2007). "Control Chart for the Coefficient of Variation". Journal of Quality Technology. Vol. 39.No 2.

FORKMAN, J.; VERRILL, S. (2008) "The distribution of Mc Kay's approximation for the Coefficient of Variation". Statistics and Probability Letters.

HENDRICKS , W.; ROBEY, W. (1936). "The sampling distribution of the coefficient of variation." Annals Math. Stat. Vol.7,129-132.

IGLEWICZ, B ; MYERS, R.(1970). "Comparisons of approximations to the percentage points of the sample Coefficient of Variation". Technometrics,12,166-169

UMPHREY, G. J. (1983) "A comment on McKay's approximation for the Coefficient of Variation. Comm.Stat.-Sim & Comp.12(5),629-635.

WEI LIU et al.(2006) "Exact confidence bounds for the Coefficient of Variation of a Normally Distributed Population". International Journal of Statistics and Systems. Vol. 1, N°1, 81-86.