

Control estadístico de la media con datos censurados a intervalo

Statistical control of the mean in presence of interval-censored data

Romario Conto López ^{a*}
Nelfi González Álvarez^b

Fecha de Recepción: 12 - may. - 2016.

Fecha de Aceptación: 12 - sep. - 2017.

Resumen

En ocasiones, durante el monitoreo de procesos no es posible realizar mediciones exactas sobre las unidades observadas, por lo que se deberá trabajar con observaciones censuradas y, debido a esto, no sería apropiada una aplicación directa de las cartas de control tipo Shewhart tradicionales. Teniendo en cuenta esta situación, se debe recurrir a cartas de control que consideren el tipo de censura presente y una distribución adecuada para los datos. Para el caso de monitoreo sobre datos con censura a intervalo, bajo un esquema de inspección donde las unidades buenas son las que quedan censuradas y por ende se presenta alta censura en estado estable del proceso, se proponen tres cartas de control para \bar{X} , donde en cada una se imputa la censura de manera diferente, asumiendo distribución normal para la variable cuya media se controla, y censura tipo I. En la primera carta, cada observación censurada se reemplaza por el valor esperado condicional (CEV), en la segunda, se reemplaza por el valor del límite superior de censura, y en la tercera, se reemplaza por un valor aleatorio de una uniforme dentro del intervalo de censura. Mediante simulación se estudia la potencia en la detección de desviaciones de la media para las cartas propuestas y se compara con la carta Shewhart tradicional que ignora la censura.

Palabras clave: Cartas de control, censura, monitoreo de procesos.

Abstract

Occasionally, during the statistical process monitoring it is not possible to make measurements with precision on the observed units; therefore, it should be necessary to work with censored observations. Consequently, a direct application of the traditional Shewhart control charts would not be appropriate. Under these circumstances, it should be implemented control charts that include both the censoring involved and a suitable distribution for the resulting data. Under an inspection scheme, where no defective units are censored and a high censoring is presented on the stable state of the process, three control charts for \bar{X} are proposed, in the case of monitoring interval-censored data. Each of them has a different way to impute the censored data. In the first one, each censored observation is replaced by its conditional expected value (CEV); in the second one, the censored data is replaced by the upper limit of the interval which the censored data belongs to, and in the third one, it is replaced by a random value of a uniform distribution within the censoring interval, considering a typical distribution and censoring type I. The power for detecting deviations from the mean for the proposed charts is compared by means of simulation with the Shewhart chart that ignores censoring.

Keywords: Censoring, control charts, monitoring process.

a Universidad Nacional de Colombia, Escuela de Estadística.

*Correo electrónico: racontol@unal.edu.co

b Universidad Nacional de Colombia, Escuela de Estadística.

*Correo electrónico: ngonzale@unal.edu.co

1. INTRODUCCIÓN

En ciertos casos el monitoreo de procesos trabaja con datos censurados, situación en la cual no se conoce el valor exacto de una característica de calidad monitoreada, básicamente debido a que durante el estudio, por posibles limitaciones de tiempo, costos, o en la precisión del instrumento de medición, no se logra conocer el valor exacto de la característica de calidad y la única información que se obtiene es que esta se encuentra dentro de algún intervalo. En la literatura del control estadístico de procesos se ha abordado la censura por diferentes autores. Por ejemplo [1] estudian el caso de censura a derecha durante el monitoreo de procesos, centrándose en censura a derecha tipo I, con una proporción de censura mayor al 50 %, y proponen una carta de control \bar{X} basada en el valor esperado condicional (CEV), considerando una distribución normal para la variable de interés bajo control. Así mismo [2] proponen una carta \bar{X} CEV con censura a derecha considerando una distribución Weibull para los datos. [3] plantean el caso de censura tipo II y construyen una carta con el fin de monitorear el parámetro de forma en la distribución Weibull. [4] desarrollan una carta CUSUM basada en la razón de verosimilitud para el monitoreo de los parámetros de escala y de forma de la distribución Log-logística con censura a derecha. [5] exponen el uso del CEV convergente para imputar la censura, centrándose en caso de censura a derecha con distribución Weibull. [6] contruyen una carta para monitorear la media del proceso considerando censura a derecha y una distribución Rayleigh. Los trabajos antes mencionados, en relación al monitoreo de media del proceso, mejoran el rendimiento de la carta tipo Shewhart tradicional con censura alta (mayor al 90 %), bajo la distribución dada, pues estas cartas mostraron ser más potentes en la detección de desviaciones de la media del proceso que la tradicional.

[1] exponen una situación en la cual son usadas barras mecanizadas como galgas, con un diámetro determinado por los límites de especificaciones de producto, para controlar el diámetro de ciertas arandelas involucradas en la producción de toma - corrientes. Si la barra con mayor diámetro entra en el orificio de las arandelas, entonces se ha excedido el límite de especificación superior, y si la barra más pequeña no entra

en el orificio, el diámetro de esta es menor que la especificación inferior. Para las unidades que no están dentro de especificaciones se les mide el diámetro, en tanto, que las unidades dentro de especificaciones quedan censuradas en el intervalo definido por los límites de especificación. Este tipo de censura es conocida como censura tipo I a intervalo. El interés del presente artículo es proponer una metodología para la construcción de una carta de control tipo Shewhart que permita monitorear la media del proceso cuando se tiene censura tipo I a intervalo y la variable a monitorear es $X \sim N(\mu, \sigma^2)$; también se estudia la potencia de las cartas propuestas en la detección de desviaciones en la media de X , en la fase II de control, en comparación con la carta tipo Shewhart tradicional, con el fin de determinar si las mejoras halladas por otros autores en el tema de la censura en control estadístico de procesos con datos censurados, se mantienen para el caso de censura a intervalo. Cabe notar que bajo el esquema de monitoreo planteado, las unidades buenas son las que quedan censuradas y por tanto se espera una alta censura en estado estable, pues de lo contrario el proceso no estaría en control y no sería un proceso capaz ni se podría seguir el esquema de monitoreo descrito. También se pretende determinar, cuál mecanismo de imputación puede ser más apropiado, entre aquel basado en el valor esperado condicional (CEV), usando el límite superior del intervalo de censura, o bien, usando un valor aleatorio de una uniforme en el intervalo $[a, b]$ de especificación.

En la sección 2 se presentarán las tres cartas propuestas que permiten abordar la problemática de la censura. En la sección 3 se estudia la potencia de cada una en la detección de cambios en la media del proceso en comparación con la carta tipo Shewhart. Finalmente se presenta un caso simulado donde se puede evidenciar la eficacia de las cartas y finalmente concluir sobre la mejor metodología a seguir.

2. CARTAS \bar{X} EN DATOS CON CENSURA A INTERVALO

En esta sección se procede a desarrollar tres cartas de control \bar{X} para tratar la falta de información resultante por la censura a intervalo, la cual se presenta cuando una observación sobre

un característica de calidad X , queda dentro del intervalo de censura $[a, b]$, siendo a y b los límites de especificación para esta característica y por tanto, no es observada su medida exacta, solo se sabe que está dentro de dicho intervalo. La idea de las cartas consiste en reemplazar cada observación censurada por un valor cercano al verdadero que tomaría la variable si hubiese sido observada. Para ello, se estudian tres mecanismos de imputación, a saber: la imputación mediante el valor esperado condicional (CEV), imputando la observación censurada con el límite superior del intervalo de censura y por último se usará el método mencionado por [7] donde se propone reemplazar la observación censurada con el valor de una uniforme dentro del intervalo de censura.

La finalidad de las cartas de control \bar{X} propuestas es monitorear y detectar cambios en la media del proceso cuando se presenta censura tipo I a intervalo en las observaciones, las cuales tendrán tanto límite inferior como límite superior de control. Para hallar los límites de control en las tres cartas se considera un proceso bajo control y que la variable aleatoria X , que representa la característica de calidad a monitorear, tiene una distribución $N(\mu, \sigma^2)$. En todos los casos, los límites serán inicialmente calculados de forma estándar considerando que $X \sim N(0,1)$ con una cantidad de subgrupos $m = 10000$, tamaños de subgrupo $n = 3,5,10,20$ y diferentes proporciones de censura p entre 0 y 1. Los límites de control requeridos en las cartas propuestas para una distribución $N(\mu, \sigma^2)$ pueden ser obtenidos como [1]:

$$LCL_{\bar{X}} = LCL * \sigma + \mu$$

$$UCL_{\bar{X}} = UCL * \sigma + \mu$$

Donde μ y σ son los parámetros del proceso bajo control y el UCL y LCL son los límites superior e inferior de control estándar obtenidos mediante simulación bajo una $N(0,1)$. Los límites estándar dependerán de la carta implementada.

2.1 Carta de Control CEV- \bar{X} con censura a Intervalo

Considerando el procedimiento desarrollado por [1] se procede a implementar una carta similar para el caso de censura a intervalo. La idea es re-

emplazar cada valor censurado en la muestra por el valor esperado condicional (CEV), para ello se define un intervalo de censura $[a, b]$ y de acuerdo a esto obtiene el CEV como:

$$E[X|a \leq X \leq b] = \mu + \sigma \left[\frac{\phi(Z_a) - \phi(Z_b)}{\Phi(Z_b) - \Phi(Z_a)} \right] = w_{a,b} \quad (2)$$

Donde $Z_a = (a - \mu)/\sigma$ y $Z_b = (b - \mu)/\sigma$. Luego,

se define la variable W como:

$$W = \begin{cases} w_{a,b} & a \leq X \leq b \quad (\text{Censurada}) \\ X & \text{En otro caso} \quad (\text{No Censurada}) \end{cases} \quad (3)$$

Los límites de control para la carta son probabilísticos y calculados mediante simulación, considerando una probabilidad de falsa alarma de 0.0027. El límite superior es hallado como el cuantil 0.99865 y el inferior como el cuantil 0.00135 de la distribución de las medias muestrales (\bar{W}_n), para un n dado, basado en 10000 simulaciones. De modo que:

$$LCL = q_{0.00135}(\bar{W}_n) \quad (4)$$

$$UCL = q_{0.99865}(\bar{W}_n) \quad (5)$$

Dichos límites dependen de la proporción de censura p y del tamaño de subgrupo racional n usado en el monitoreo del proceso. Los límites de control estándar para la carta CEV- \bar{X} se presentan en el Figura 1 luego de aplicar un suavizamiento spline, en la cual se logra observar que dichos límites son simétricos y se vuelven más estrechos a medida que el tamaño de subgrupo n aumenta.

2.2 Carta de control \bar{X} - b con censura a Intervalo

El procedimiento para construir la carta \bar{X} - b es similar a la carta CEV \bar{X} , solo cambia el método de imputación, donde cada dato censurado será reemplazado por el límite superior de censura, construyendo así la variable W (ver ecuación 6). Este método de imputación ha sido más popular en confiabilidad y análisis de supervivencia, donde los datos son tratados como si fueran

censurados a derecha debido a la disponibilidad de métodos para la inferencia con este tipo de censura. Es posible imputar con el punto medio del intervalo de censura, pero como se estudia en [8] dicho método puede llevar a estimaciones más sesgadas al implementar métodos no paramétricos y al trabajar bajo una distribución diferente a la normal. Los límites estándar para la carta $\bar{X} - b$ también son probabilísticos y son obtenidos de manera similar como se expone en las ecuaciones (4) y (5) para los (\bar{W}_n) en este caso. Dichos límites se pueden ver en la figura 2 luego de aplicar un suavizamiento spline, bajo los mismos niveles de censura y tamaños de subgrupo. En la Figura 2 se aprecia que los límites dejan de ser simétricos cuando la proporción de censura es mayor a 0.2

$$W = \begin{cases} b & a \leq X \leq b \quad (\text{Censurada}) \\ X & \text{En otro caso.} \quad (\text{No Censurada}) \end{cases} \quad (6)$$

2.3 Carta de control $\bar{X} - U$ con censura a Intervalo

La metodología de la carta $\bar{X} - U$ consiste en imputar cada observación censurada por un valor aleatorio de una uniforme dentro del intervalo de censura $[a, b]$, construyendo la variable W (ecuación 7). En este caso, los límites estándar para la carta también son obtenidos mediante las ecuaciones (4) y (5) para la variable resultante al aplicar el método de imputación de esta carta, los cuales son también probabilísticos. En la Figura 3 se tienen los límites de control estándar obtenidos por simulación para la carta $\bar{X} - U$, luego de aplicar un suavizamiento spline. En dicha figura se logra observar que estos son simétricos y constantes hasta una proporción de censura aproximada de 0.8, después de la cual se vuelven más amplios para todos los tamaños de muestra considerados.

$$W = \begin{cases} U(a, b) & a \leq X \leq b \quad (\text{Censurada}) \\ X & \text{En otro caso.} \quad (\text{No Censurada}) \end{cases} \quad (7)$$

Recapitulando, la construcción de las cartas propuestas para censura tipo I a intervalo implican los siguientes pasos, asumiendo μ y σ conocidos:

- Hallar el *UCL* y *LCL* mediante las Figuras 1, 2 ó 3 dependiendo de la carta a construir, del tamaño del subgrupo n y proporción de censura p .
- Sea X_{ij} la i -ésima observación en el j -ésimo subgrupo racional con $i = 1, 2, \dots, n$ y $j = 1, 2, \dots, m$. Para cada X_{ij} censurada en el intervalo $[a, b]$ se toma como observación el valor de w_{ij} según el método de imputación definido por la carta (ver ecuaciones 3, 6 y 7).
- Finalmente, calcularlo $\bar{W}_j = \frac{1}{n} \sum_{i=1}^n w_{ij}$ (promedio de subgrupos) y graficar contra límites calculados. Si hay puntos por fuera de límites se estudian las causas asignables que lo generen, permitiendo así mejorar la calidad del proceso.

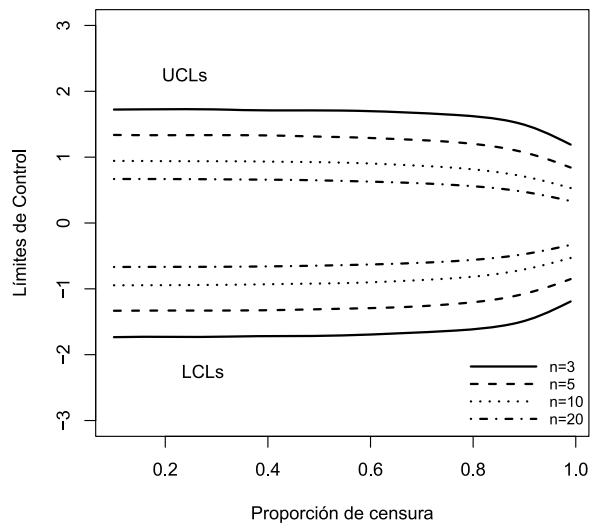


Figura 1. Límites de control estándar para la Carta CEV- \bar{X}

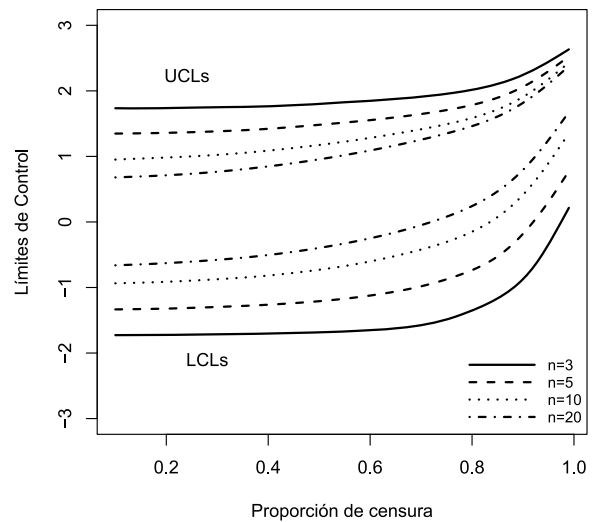


Figura 2. Límites de control estándar para la Carta $\bar{X} - b$.

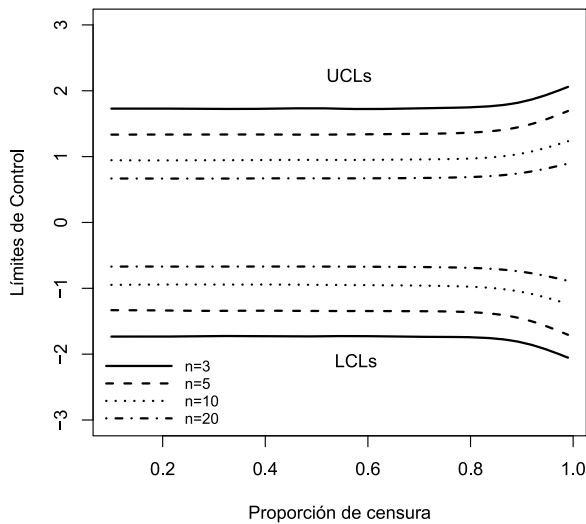


Figura 3. Límites de control estándar para la Carta \bar{X} -U.

2.4 Potencia comparativa de las cartas \bar{X} propuestas y la carta tipo Shewhart

En la práctica, la carta \bar{X} tipo Shewhart es la más usada para monitorear la media del proceso, por tanto, resulta de interés comparar la potencia de esta con las tres cartas propuestas ante censura a intervalo. Para poder realizar la comparación todas las cartas deben tener la misma probabilidad de falsa alarma cuando el proceso esté en control, la cual se fija en 0.0027. La carta \bar{X} tradicional con límites $\bar{X} \pm 3\hat{\sigma}/\sqrt{n}$ ignora la censura y la asimetría presente en las observaciones censuradas, registrando el valor del punto medio del intervalo de especificación como si fuese el valor observado para las unidades que quedan censuradas. Luego, si se usan los límites 3σ la probabilidad de falsa alarma será diferente a la deseada. Para construir la carta tipo Shewhart tradicional los límites son calculados como $\mu \pm k\sigma/\sqrt{n}$ donde el valor de k se busca tal que garantice la probabilidad de falsa alarma deseada. La potencia de cada carta es hallada como la probabilidad de detectar cambios en la media del proceso dado que efectivamente se presentó una variación, es decir, se calcula $P(LCL \leq \bar{W}_n \leq UCL | X \sim N(\mu + \delta, \sigma^2))$ con δ el cambio en la media, donde \bar{W}_n son las medias por subgrupo de tamaño n de la variable

W obtenidas luego de implementar uno de los mecanismos de imputación.

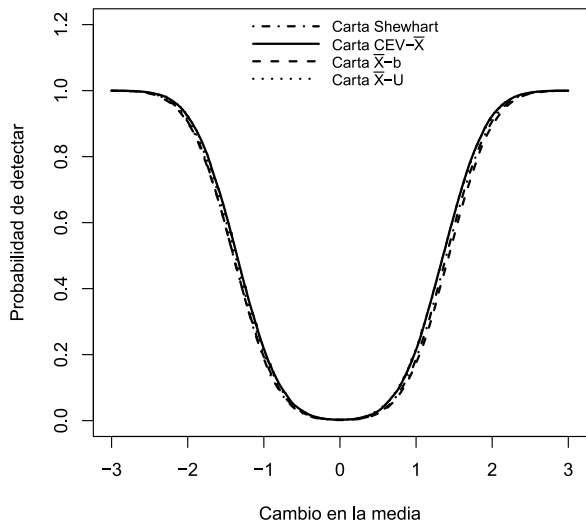


Figura 4. Potencia comparativa entre la carta Shewhart y las propuestas con $n = 5, p = 0.5$

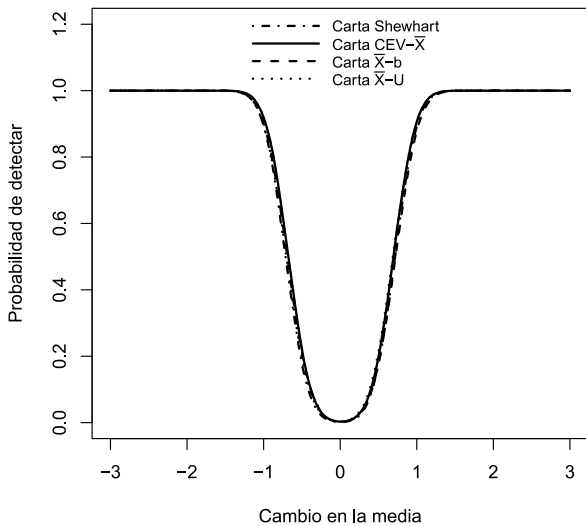


Figura 5. Potencia comparativa entre la carta Shewhart y las propuestas con $n = 20, p = 0.5$.

En la Figura 4 se puede observar que para una proporción de censura del 50 % todas las cartas tienen una potencia muy similar para un tamaño de muestra pequeño $n = 5$; algo similar sucede en la figura 5 para un tamaño de muestra mayor, $n = 20$, donde la potencia es prácticamente igual; por lo que cualquiera de estas cartas podría ser implementada.

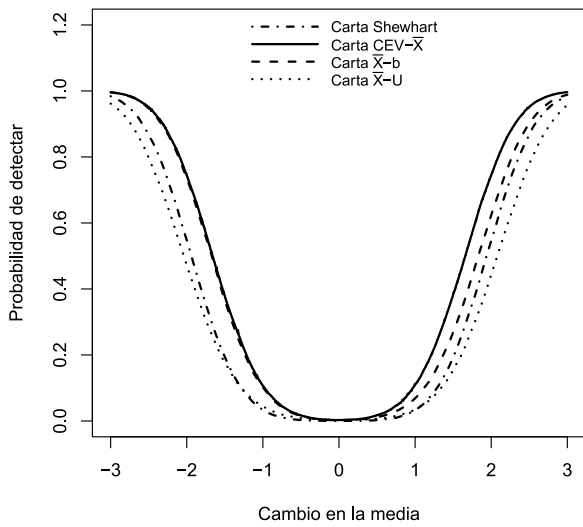


Figura 6. Potencia comparativa entre la carta Shewhart y las propuestas con $n = 5$, $p = 0.95$

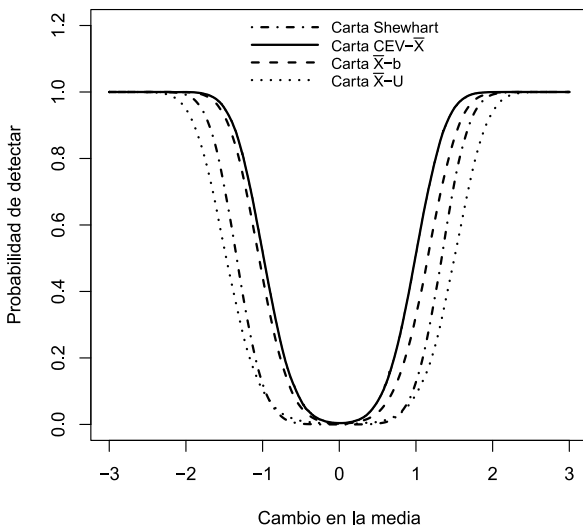


Figura 7. Potencia comparativa entre la carta Shewhart y las propuestas. con $n = 20$, $p = 0.95$.

Cuando se tiene una proporción de censura del 95%, la Figura 6 muestra que en la detección de decrecimientos en la media, las cartas CEV - \bar{X} y $\bar{X} - b$ son las más potentes, con igual probabilidad de detectar. Mientras que en la detección de incrementos en la media, la carta CEV - \bar{X} es aquella con mayor potencia, incluso respecto a la carta $\bar{X} - b$ para un tamaño de muestra $n = 5$. En la Figura 7 se observa que la carta CEV - \bar{X} es la más potente en la detección tanto de crecimientos como decrecimientos en la media, seguida por la carta $\bar{X} - b$, la cual muestra ser menos sensible en la detección de crecimientos y un poco menos potente en la detección de decrecimientos; esto debido que al imputar con el límite superior del

intervalo de censura, los límites de la carta estarán orientados a monitorear valores cercanos a dicho límite por ende más fácilmente detectarán decrecimientos que crecimientos. Las Figuras 6 y 7 muestran que la potencia de la carta Shewhart es menor respecto a las cartas CEV - \bar{X} y $\bar{X} - b$, por tanto ambas cartas propuestas mejoran el rendimiento de la Shewhart cuando se presenta censura a intervalo para ambos tamaños de muestra; también se obtiene que la potencia de la carta $\bar{X} - U$ es la más baja respecto a todas las demás con alta censura, incluso respecto a la tradicional, siendo la menos recomendable en este caso.

3. CASO SIMULADO

Con el fin de observar el desempeño de las cartas propuestas, se procede a calcular los límites de control para cada una de éstas considerando distribución normal con media $\mu = 10$, desviación estándar $\sigma = 2.5$, e intervalo de censura o límites de especificación [5.1,14.9]. En este caso se considera que los parámetros son conocidos ya que se está en fase II de monitoreo. Los límites con parámetros conocidos para las tres cartas, están determinados en la Ec. (1), donde LCL y UCL son los límites estándar dados en las Figuras 1, 2 y 3 para la proporción de censura $p = 0.95$ y $n = 5$. Por tanto,

Tabla 1. Límites de control para cada carta con $n = 5$, $\mu = 10$, $\sigma = 2.5$ y $p = 0.95$.

Carta	LCL	UCL
1 $CEV - \bar{X}$	7.51	12.48
2 $\bar{X} - b$	10.41	15.66
3 $\bar{X} - U$	6.23	13.92

Se simularon 40 subgrupos de tamaño $n = 5$ provenientes de una distribución normal con media $\mu = 10$, desviación estándar $\sigma = 2.5$ y una proporción de censura $p = 0.95$, de los cuales, los primeros $m_1 = 30$ subgrupos fueron generados en control. Cada observación dentro de límite de especificación es imputada por un valor determinado dependiendo de la carta usada. Los $m_2 = 10$ subgrupos restantes fueron simulados con corrimiento gradual en la media de 0.5σ cada dos subgrupos, generando que los últimos dos presenten un corrimiento de 2.5σ . Para la construcción de las cartas se toma en consideración los 40

subgrupos.

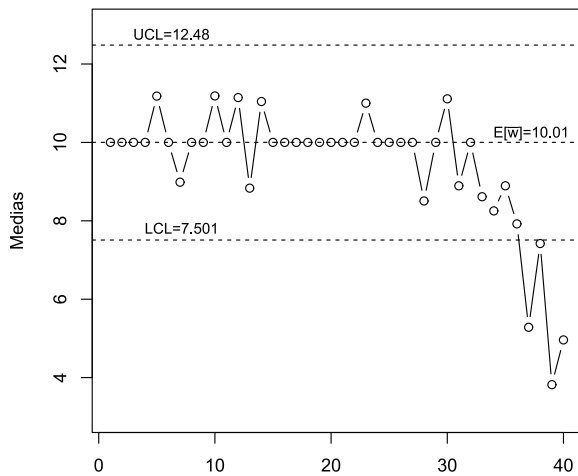


Figura 8. Carta CEV \bar{X} aplicada a caso simulado con cambio gradual en la media.

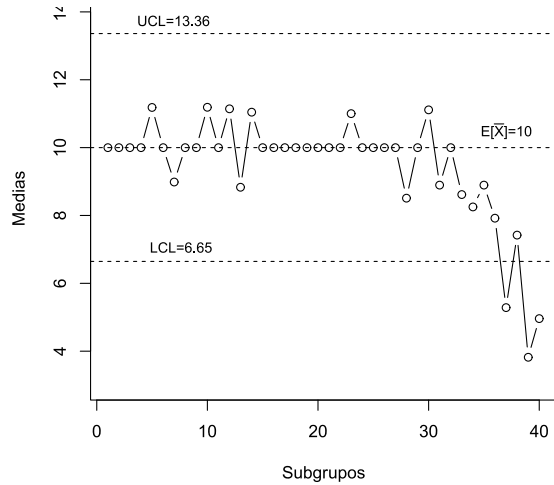


Figura 11. Carta \bar{X} Shewhart aplicada a caso simulado con cambio gradual en la media.

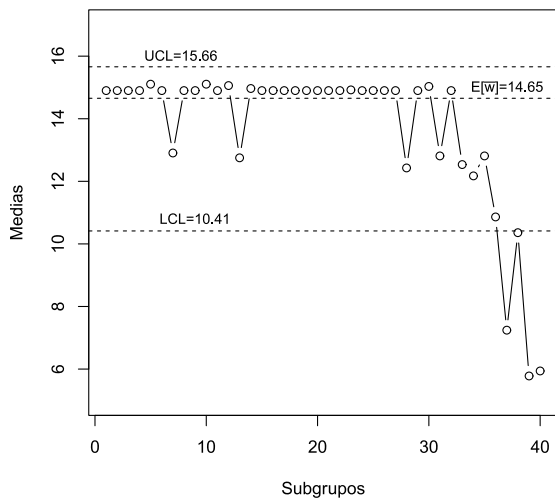


Figura 9. Carta $\bar{X}-b$ aplicada a caso simulado con cambio gradual en la media.

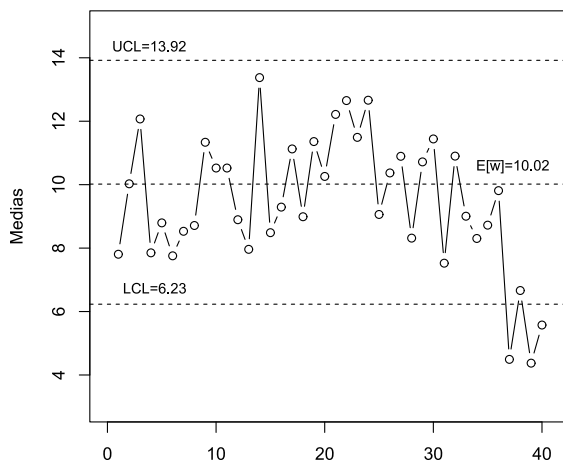


Figura 10. Carta $\bar{X}-U$ aplicada a caso simulado con cambio gradual en la media.

En las figuras 8, 9, 10 y 11 se pueden observar las cartas de control propuestas y la carta Shewhart tradicional para los subgrupos simulados. En dichas figuras se puede ver que para “grandes” corrimientos en la media todas las cartas logran detectar; pero son las cartas CEV- \bar{X} y $\bar{X}-b$ las que detectan más rápido que el proceso está fuera de control, ya que como se muestra en las figuras 8 y 9, detectan punto por fuera de límites a partir del subgrupo 37, con un decrecimiento de 2σ en la media, e incluso para el subgrupo 36, con un corrimiento de 1.5σ ya se está bastante cerca del límite inferior para estas dos cartas, lo cual no se observó en la carta $\bar{X}-U$ (figura 10) ni en la carta tradicional (figura 11). Adicionalmente, en las cartas CEV- \bar{X} y $\bar{X}-b$ hay un cambio en el patrón de comportamiento con tendencia de decrecimiento sistemático, cosa que no se evidencia en las otras dos cartas. Por tanto, las cartas $\bar{X}-U$ y la tradicional no reflejan eficazmente que el proceso se ha salido de control.

4. CONCLUSIONES

Cuando se desea monitorear la media de un proceso productivo y se tiene un conjunto de observaciones registradas bajo censura a intervalo tipo I con una alta proporción de censura, la carta de control CEV- \bar{X} , considerando normalidad en los datos, puede ser aplicada, pues demuestra ser más potente para detectar cambios en la media del proceso que la carta tipo Shewhart tradicional, tanto en los casos con proporción de censura del 50 y del 95% y tamaños de muestra pequeños y grandes. Con alta proporción de censura

(95%), la carta CEV- \bar{X} es un poco más sensible a decrecimientos en la media que la carta $\bar{X}-b$ y aún más sensible cuando hay incrementos. La carta $\bar{X}-U$ es más insensible a movimientos en la media, pues imputa con un valor al azar. Estas diferencias con alta censura son el escenario a destacar, pues es lo esperado para un esquema de monitoreo donde las unidades aceptadas son las censuradas. Cabe resaltar que a medida que la proporción de censura aumenta todas las cartas van perdiendo potencia, pero es la carta CEV- \bar{X} quien logra mejores resultados. Las cartas de control propuestas no solo permiten ver si el proceso está en control, si no que adicionalmente ayudan a determinar si hay corrimientos en la media que puedan generar una señal futura, por lo cual se prefieren en lugar de las cartas por atributos que monitorean solo si la variable está o no está dentro de un intervalo de especificación.

Adicionalmente, se tiene que las cartas propuestas son para fase II de control, donde se asumen parámetros conocidos, y si estos no se conocen deben ser estimados basados en una muestra inicial, en la cual es necesario obtener observaciones exactas para poder tener información con la cual estimar con cierto nivel de precisión y estudiar el efecto de la estimación de parámetros en el desempeño de las cartas propuestas.

AGRADECIMIENTOS

A la convocatoria “Programa Nacional de apoyo a estudiantes de posgrado para el fortalecimiento de la investigación, creación e innovación de la Universidad Nacional de Colombia 2013-2015”, la cual brindó los recursos económicos para llevar a término el trabajo plasmado en este artículo.

REFERENCIAS

- [1] H. S. Steiner y J. R. Mackay, “Monitoring processes with highly censored data”, *Journal of Quality Technology*, vol. 32, no. 3, pp. 199-208, 2000.
- [2] J. Vargas y T. Montaña, “Carta de control CEV X para distribuciones Weibull con datos censurados”, *Revista Colombiana de Estadística*, vol. 28, no. 2, pp. 125-139, 2005.
- [3] Y. Chan, B. Han y F. Pascual “Monitoring the Weibull shape parameter with Type II censored data”, *Quality and Reliability Engineering International*, vol. 31, no. 5, pp. 741-760, 2015.
- [4] B. Sadeghpour y M. Taghizadeh, “Optimal cusum control chart for censored reliability data with Log-logistic distribution”, *Computational Methods in Science and Technology*, vol. 21, no. 4, pp. 221-227, 2015.
- [5] Y. He, L. Wang, Z. He y Y. Wei, “Product reliability oriented design scheme of control chart based on the convergent CEV for censored characteristics”, *Mathematical Problems in Engineering*, vol 2015, Article ID 128491, 11 pages, 2015.
- [6] S. M. Raza, M. Moeen, M. Azad y A. F. Siddiqi, “Shewhart control charts for Rayleigh distribution in the presence of type I censored data”, *Journal of ISOSS*, vol. 2, no. 2, pp. 210-217, 2016.
- [7] C. Hsu, M. G. Taylor, S. Murray y D. Comenges, “Multiple imputation for interval censored data with auxiliary variables”, *Statistics in Medicine*, vol. 25, pp. 769-781, 2006.
- [8] A. Zyoud, F. A. Elfaki y M. Hrairi, “Non-parametric estimate based in imputations techniques for interval and partly interval censored data”, *Science International (Lahore)*, vol. 28, no. 2, pp. 879-884, 2016.