

ESTIMACIÓN DE ÍNDICES DE CAPACIDAD DE PROCESOS USANDO LA DISTRIBUCIÓN GENERALIZADA DE PARETO

ESTIMATION OF CAPABILITY PROCESS INDICES USING THE GENERALIZED PARETO DISTRIBUTION

Martha Rosa Valdiviezo Márquez¹, José Simón Fermín²

¹ Universidad de Oriente. Escuela de Administración. Departamento de Contaduría. Código Postal 6101. Cumaná-Venezuela.

² Instituto Universitario de Tecnología. Código Postal 6101. simon.fermin@gmail.com

RESUMEN

Los índices de capacidad de procesos (ICP) son un medio altamente efectivo de determinar la calidad del producto y desempeño del proceso. Entre muchos índices de capacidad de procesos desarrollados, C_p , C_{pk} , C_{pm} y C_{pmk} son los cuatro índices más populares bajo procesos distribuidos normalmente. Sin embargo, cuando estos índices tradicionales son utilizados para evaluar un proceso distribuido no normalmente a menudo guía a resultados inexactos. Por esto, los ICP basado tanto en el método de percentiles de Clements como en el método de percentiles de Burr fueron propuestos para superar esta deficiencia bajo procesos distribuidos no normalmente. Por esta razón, el objetivo de este trabajo es determinar el desempeño y la confiabilidad del método de bootstrap para estimar los intervalos de confianza para los índices basados en la técnica de Percentiles de Clements, Percentiles de Burr y el índice normal usando la Distribución Pareto Generalizada con dos parámetros. Posteriormente se comparan entre sí los intervalos de confianza Estándar Bootstrap, Percentil Bootstrap y Percentil Sesgo Corregido Bootstrap. Se realizó una serie de simulaciones usando la distribución Pareto Generalizada con diferentes condiciones, resultando en forma general que el índice estimado con el método de percentil de Burr es mejor estimador en cuanto al porcentaje de cubrimiento y ancho promedio. Por otro lado, cuando el valor de $C_{pu} = 0,50$, el intervalo percentil bootstrap es en promedio, mejor estimador que los intervalos estándar bootstrap y percentil sesgo corregido. Para diferentes condiciones ($C_{pu} = 1,0$ y $1,5$), el método de percentil sesgo-correcto estima mejor los intervalos de confianza según los métodos de percentiles de Clements y Burr.

Palabras claves: Distribución generalizada de Pareto, estimación, índices de capacidad.

ABSTRACT

The processes capability indices (PCI) are highly effective in determining the quality of the product and process performance. Among many indices developed process capability, C_p , C_{pk} , C_{pm} and C_{pmk} are the four most popular indices under normally distributed processes. However, when these traditional indices are used to assess a non-normally distributed process often leads to inaccurate results. For this, ICPs based on both Clements percentiles of method and the Burr method of percentiles were proposed to overcome this deficiency under processes not normally distributed. Therefore, the aim of this study is to determine the performance and reliability of the Bootstrap method to estimate confidence intervals for the ICP technique based on Clements Percentiles, Burr Percentiles and normal C_{pk} using the Generalized Pareto Distribution with two parameters. Then are compared with one another the Standard Bootstrap confidence intervals, Percentile Bootstrap and Biased Corrected Percentile Bootstrap. A series of simulations was conducted using the Generalized Pareto Distribution with different conditions, generally resulting in the index with Burr percentile method is a better estimator in terms of percentage of coverage and average width. On the other hand, when the value of $C_{pu} = 0,50$ bootstrap percentile interval estimator is on average better than standard bootstrap and

Biased Corrected Percentile Bootstrap intervals. For different conditions ($C_{pu} = 1,0$ y $1,5$) the biased corrected percentile bootstrap method is better to estimate confidence interval of PCI according to the percentile methods of Clements and particularly of Burr.

Keywords: Generalized Pareto distribution, estimation, capability indices.

INTRODUCCIÓN

Los índices de capacidad de procesos, se utilizan para el monitoreo de la capacidad de un proceso, estos proporcionan medidas numéricas de la capacidad y mejoramiento de dicho proceso. Los índices más comúnmente utilizados son el C_p , C_{pk} , C_{pm} y C_{pmk} .

Han (2006), expone una forma más generalizada denominada forma de superestructura. Esta forma es la siguiente.

$$C_p(u,v) = \frac{d-u |\mu-m|}{3\sqrt{\sigma^2+v(\mu-T)^2}} \quad (1)$$

con, $u \geq 0$, $v \geq 0$, y donde, μ es la media del proceso, σ es la desviación estándar del proceso,

T es un valor target, $d = \frac{(U-L)}{2}$, U y L son los límites de especificación superior e inferior

respectivamente y $m = \frac{(U+L)}{2}$. De manera más específica se tiene lo siguiente.

$$C_p(0,0) = C_p = \frac{U-L}{6\sigma} \quad (2)$$

$$C_p(1,0) = C_{pk} = \min \left\{ \frac{U-\mu}{3\sigma}, \frac{\mu-L}{3\sigma} \right\} \quad (3)$$

Un supuesto importante en la aplicación e interpretación de dichos índices es la condición de normalidad de la salida del proceso. Si esta distribución fundamental es no-normal, entonces los enunciados acerca de la fracción no-conforme esperada pueden generar resultados erróneos. Los índices de capacidad de procesos C_p, C_{pk} , son discutidos por Kane (1986) o en Kotz & Johnson (1993).

En la industria, existen muchos procesos de manufacturas donde la variabilidad del proceso no se puede describir por una distribución normal. Las distribuciones asimétricas acotadas por un lado, son bastante común en la industria. Para este tipo de procesos no normales, no se debería usar las mismas fórmulas para calcular los índices de capacidad de procesos, porque los resultados darían resultados engañosos acerca del desempeño del proceso.

En la literatura se encuentran diversos métodos para analizar datos no normales para el cálculo de índices de capacidad de procesos. Un enfoque para tratar con datos no normales, es transformarlos usando una función matemática, de tal manera que los datos resultantes sean normales. Otro punto de vista para el estudio de datos provenientes de procesos no normales,

según Chou, Polanski & Mason (1998), Pal (2005), “es ajustar una distribución no normal estándar (Weibull, log-normal, Gamma) o una familia de curvas de distribuciones (Pearson y Johnson) de los datos del proceso y entonces estimar los puntos percentiles provenientes de la distribución ajustada, para luego utilizarlos en el cálculo de los índices de capacidad de procesos”.

En complemento con esto, Clements (1989), propuso un índice de capacidad de procesos generalizado, y lo denota “ C_{pk}^* ”. El procedimiento consiste en calcular los índices de capacidad de procesos como una modificación de los índices de capacidad de procesos normales, utilizando los valores de puntos percentiles no normales y considerando la mediana como medida central.

En este trabajo se considera la Distribución Pareto Generalizada con dos parámetros (DPG2), para evaluar el comportamiento de los índices de capacidad de procesos no normales. Se estimarán estos índices tomando en cuenta el método de Clements, el de Burr y el método cuando se considera normal la distribución de los datos. Para el logro de este objetivo se toman en cuenta estimadores puntuales y por intervalos bootstrap (percentil bootstrap (PB), estándar bootstrap (EB) y percentil sesgo-correctado bootstrap (PSCB)), generados mediante simulaciones y considerando una muestra cuya distribución es PG2.

MATERIALES Y MÉTODOS

En este estudio se realizan diversas simulaciones para la obtención de intervalos de confianzas bootstrap de 95% con el objetivo de estimar los índices de capacidad de procesos mediante los métodos de Burr, Clements y Normal. Se utiliza una muestra real de tamaño 198 proveniente de una distribución Pareto Generalizada la cual se describe a continuación.

Distribución Pareto Generalizada

La Distribución de Pareto, fue propuesta en 1897 por Pareto, como un “modelo para distribuciones de ingreso”; actualmente puede ser utilizada para representar otras distribuciones, de hecho es una combinación de distribución exponencial y gamma. Juega un papel muy importante en el estudio sobre el tamaño de poblaciones en ciudades, acontecimientos que involucren riquezas naturales, riesgo en seguros y fallas comerciales.

Pickands (1975), la denominó “Distribución de Pareto Generalizada (Generalized Pareto Distribution, GPD)”, luego fue utilizada por Davinson (1984) y Van Montfort & Witter (1985). Desde ese entonces, su aplicabilidad se ha visto fuertemente relacionada en análisis para valores extremos, modelamiento de fenómenos socioeconómicos, procesos físicos y biológicos (Saksena & Johnson (1984)), análisis de medio ambiente, distribución para tiempos de fallas; en fin, situaciones en las cuales la distribución exponencial se puede utilizar pero se requiere algún estudio de robustez en alternativa de colas.

Una definición de esta distribución sería la siguiente (Hosking 1987). “La Distribución Pareto

Generalizada (DPG2), es la distribución de una variable aleatoria X , definida como
$$X = \frac{\alpha(1 - e^{-\alpha Y})}{\beta}$$

, donde Y es una variable aleatoria con distribución exponencial estándar, α y β son parámetros. La distribución de X se dice que es una DPG2, si la función de distribución está definida de la siguiente manera,

$$F(x) = \begin{cases} 1 - \left(1 - \frac{kx}{\alpha}\right)^{1/k} & \text{si } k \neq 0 \\ 1 - \exp\left(-\frac{x}{\alpha}\right) & \text{si } k = 0 \end{cases} \quad (4)$$

Así mismo, la fdp está definida de la siguiente manera;

$$f(x) = \begin{cases} \alpha^{-1} \left(1 - \frac{kx}{\alpha}\right)^{\frac{1-k}{k}} & \text{si } k \neq 0 \\ \alpha^{-1} \exp\left(-\frac{x}{\alpha}\right) & \text{si } k = 0 \end{cases} \quad (5)$$

donde, α el parámetro de escala y k el parámetro de forma. El rango de x está dividido de la siguiente manera, para $k \leq 0$, se tiene $0 \leq x < \infty$ y para $\alpha > 0$, se tiene, $0 \leq x < \alpha/k$. Cuando $k=0$, se obtiene la distribución exponencial con media α y cuando $k=1$ se tiene la distribución uniforme sobre $[0, \alpha]$. Para $k < 0$, esta distribución se denomina Pareto. Para $k > 1/2$ la distribución PG presenta finitos puntos finales para $f(x) > 0$.

Por otro lado, la función de estimación de los cuantiles de la DPG2 está dada de la siguiente manera;

$$x(F) = \begin{cases} \frac{\alpha \{1 - (1-F)^k\}}{k} & \text{si } k \neq 0 \\ -\alpha \log(1-F) & \text{si } k = 0 \end{cases} \quad (6)$$

Un estimador $\hat{x}(F)$ se define sustituyendo los valores de $\hat{\alpha}$ y \hat{k} para los valores obtenidos en la ecuación (6).

Finalmente, las propiedades de la DPG2 y los estimadores de sus parámetros se presentan a continuación.

1. Media, $E[x] = \frac{\alpha}{1+k}$

2. Varianza, $\text{Var}(x) = \frac{\alpha^2}{(1+k)^2(1+2k)}$

3. Coeficiente de asimetría, $\gamma(x) = \frac{2(1-k)(1+2k)^{3/2}}{(1+3k)}$

4. Coeficiente de kurtosis; $\kappa(x) = \frac{3(1+2k)(3-k+2k^2)}{(1+3k)(1+4k)} - 3$

5. El estimador de momentos de α y k , son, $\hat{\alpha} = \frac{1}{2} \bar{x} \left(\frac{\bar{x}^2}{s^2} + 1 \right)$ y $\hat{k} = \frac{1}{2} \left(\frac{\bar{x}^2}{s^2} - 1 \right)$ donde, \bar{x} y s^2 son la media y la varianza muestral.

Índices de capacidad de procesos no normales

Un índice de capacidad, es una función de los parámetros de la distribución y las especificaciones establecida por la empresa. La idea principal es que el proceso productivo sigue una distribución normal con media μ y desviación estándar σ . Este supuesto en la industria no siempre puede ser apropiado ya que existen muchos procesos de manufacturas donde la variabilidad del proceso no se puede describir por una distribución normal. Muchos autores han estudiado el efecto causado por la condición de no-normalidad de los datos, en el comportamiento de los índices de capacidad de procesos.

En este trabajo se consideran dos métodos para estimar los índices de capacidad de procesos no normales que se describen en la siguiente sección y solo para efecto de comparar se considera también la metodología de los índices de capacidad de procesos normales.

Método de percentiles de Clements

Este método es muy utilizado en la industria, Clements (1989), lo propuso “para determinar los percentiles basado en la familia de distribuciones de Pearson y para determinar los índices de capacidad de procesos”. Reemplaza en la ecuación (2), el valor 6σ por la longitud del intervalo entre el límite superior e inferior de los puntos percentiles 0,135 y 99,865 de una distribución X, esto es respectivamente $U_p - L_p$. Se define el C_p de la siguiente manera.

$$CC_p = \frac{U - L}{U_p - L_p} \quad (7)$$

donde, U_p es el percentil 99,865 y L_p es el percentil 0,135 de la distribución considerada y es presentado por Gruska, Mirkhani, & Lamberson (1989); los valores U y L son los límites de especificación superior e inferior respectivamente.

Para el índice tradicional C_{pk} , la media del proceso μ , es estimada por la mediana M, y los valores de 3σ son estimados por $M - L_p$ y $U_p - M$ respectivamente. Así se tienen las siguientes expresiones,

$$CC_{pi} = \frac{M - L}{M - L_p} \quad (8)$$

$$CC_{ps} = \frac{U - M}{U_p - M} \quad (9)$$

Finalmente,

$$CC_{pk} = \text{Min}(CC_{ps}, CC_{pi}) \quad (10)$$

Este método utiliza el estimador clásico de asimetría y kurtosis los cuales están basados en el tercer y cuarto momento de la distribución.

Tanto el tercer como el cuarto momento se utilizan en una adecuada distribución de Pearson, además los puntos percentiles superior e inferior también se obtienen con esta distribución utilizando un conjunto de datos.

Método de percentiles de Burr

Burr (1942), propuso una distribución llamada "Distribución Burr XII" para obtener los percentiles requeridos de una variable X. La función de densidad de probabilidad de una variable X con distribución Burr XII está dada de la siguiente manera.

$$f(x|c,k) = \begin{cases} ckx^{c-1} (1+x^c)^{-(k+1)} & \text{si } x \geq 0, c \geq 1, k \geq 1 \\ 0 & \text{si } x < 0 \end{cases} \quad (11)$$

donde, c y k representan los coeficientes de asimetría y kurtosis de la distribución Burr. Por otro lado, la función de distribución acumulada está dada por.

$$F(x|c,k) = \begin{cases} 1 - (1+x^c)^{-k} & \text{si } x \geq 0, c \geq 1, k \geq 1 \\ 0 & \text{si } x < 0 \end{cases} \quad (12)$$

Burr (1942) y (1973), logró tabular los valores esperados de la desviación estándar, el coeficiente de asimetría y kurtosis de la distribución Burr XII para diferentes combinaciones de c y k. Estos valores se encuentran tabulados en Liu & Chen (2006).

En este método también se utilizan el tercer y cuarto momento (s_3 y s_4), de una distribución empírica, los cuales están dados por las siguientes expresiones matemáticas, respectivamente.

$$s_3 = \frac{n}{(n-1)(n-2)} \sum_i \left(\frac{x_i - \bar{x}}{s} \right)^3$$

$$s_4 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_i \left(\frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n+1)^2}{(n-2)(n-3)}$$

donde, \bar{x} es la media muestral y s es la desviación estándar muestral.

Finalmente, Liu & Chen (2006), introdujeron una modificación del método de Percentiles de Clements el cual utiliza los "puntos percentiles en una familia de curvas de Pearson". En este método se sustituyen estos puntos por percentiles apropiados obtenidos por la distribución de Burr. Este método se resume en los siguientes pasos.

Paso 1: Se estiman la media (\bar{x}), desviación estándar, el coeficiente de asimetría (s_3) y la kurtosis (s_4) del conjunto de datos con distribución X dada.

Paso 2: Se seleccionan los parámetros c y k basados en los valores estimados de asimetría y kurtosis obtenidos en el paso anterior usando la distribución Burr XII, los cuales están tabulados en Liu y Chen (2006) y cuyas expresiones son las siguientes.

$$c = \frac{(n-2)}{\sqrt{n(n-1)}} s_3 \quad \text{y} \quad k = \frac{(n-2)(n-3)}{(n^2-1)} s_4 + 3 \frac{(n-1)}{(n+1)}$$

Paso 3: Con esta misma tabla y de acuerdo a estos parámetros c y k obtenidos en el paso anterior se obtienen los valores de los percentiles estandarizados 0,00135; 0,5 y 0,99865 de la distribución Burr XII, estos percentiles se denotan como $BZ_{0,00135}$, $BZ_{0,5}$ y $BZ_{0,99865}$ respectivamente. De manera similar están tabulados los valores percentiles que se utilizan en el método de Clements.

Paso 4: Se estiman los percentiles inferior, mediana y superior denotados como sigue, $X_{0,00135}$, $X_{0,5}$ y $X_{0,99865}$ y que además son los valores de L_p , M y U_p ; estos se calculan de la siguiente manera.

$$L_p = X_{0,00135} = \bar{x} + s_x Z_{0,00135} \quad (13)$$

$$M = X_{0,50} = \bar{x} + s_x Z_{0,50} \quad (14)$$

$$U_p = X_{0,99865} = \bar{x} + s_x Z_{0,99865} \quad (15)$$

Paso 5: Finalmente se estiman los índices de capacidad de procesos mediante las fórmulas (7)-(10).

Métodos Bootstrap

Franklin & Wasserman (1992), desarrollaron métodos bootstrap para determinar intervalos de confianza considerando distribuciones normales y no normales. Efron & Tibshirani (1986), consideraron estos tres tipos de intervalos de confianza Bootstrap, los cuales se describen mas adelante.

Para construir los intervalos de confianza Bootstrap, se consideran una muestra de tamaño n, (x_1, x_2, \dots, x_n) tomadas de un proceso y se determinan los índices de capacidad ya sean por el método tradicional (C_{pk}), el de Burr, (BC_{pk}), o el de Clements, (CC_{pk}). En función a esto, se genera una muestra bootstrap (B) de tamaño 1000, sin reemplazamiento para cada uno de los diferentes índices y luego se ordena de mayor a menor lo cual constituye la distribución de los índices respectivos.

A continuación se definen los tres diferentes intervalos Bootstrap a utilizar.

Estándar Bootstrap (EB)

Sea $\hat{C}^*(i)$, el i-ésimo estimador bootstrap ordenado del índice C^* con $i=1,2,\dots,B$. y sean,

$$\bar{C}^* = \frac{1}{B} \sum_{i=1}^B \hat{C}^*(i) \quad \text{y} \quad S_C^* = \left[\frac{\sum_{i=1}^B (\hat{C}^*(i) - \bar{C}^*)^2}{B-1} \right]^{\frac{1}{2}}$$

La cantidad S_C^* es un estimador de la desviación estándar de \hat{C} , entonces (si la distribución de \hat{C} es aproximadamente normal) el $(1-2\alpha)100\%$ intervalo de confianza Bootstrap estándar está dado por,

$$\bar{C}^* \pm z_\alpha S_C^* , \tag{16}$$

donde, z_α es el $(1-2\alpha)100$ valor del percentil de la distribución normal estándar, en nuestro caso 1,96.

Percentil Bootstrap (PB)

Para la colección ordenada considerada, el estimado $\hat{C}^*(i)$ del $(1-2\alpha)100\%$ intervalo de confianza Bootstrap percentil está dado por.

$$[\hat{C}^*(\alpha B), \hat{C}^*((1-\alpha)B)] \tag{17}$$

Percentil sesgo-correcto Bootstrap (PSCB)

Para este método puede darse el caso que la distribución Bootstrap obtenida para las B muestras pueda ser sesgada, con este método se trata de corregir este sesgo.

Primeramente, usando la distribución ordenada de \hat{C}^* , se calcula la probabilidad siguiente,

$$p_0 = \Pr[\hat{C}^* \leq \hat{c}]$$

donde, \hat{C} está por el valor del índice de capacidad estimado de la muestra original de donde se tomó la muestra Bootstrap. Se calculan los siguientes valores.

$$z_0 = \Phi^{-1}(p_0)$$

$$P_{LEI} = \Phi(2z_0 - z_\alpha)$$

$$P_{LES} = \Phi(2z_0 + z_\alpha)$$

donde, $\Phi(.)$ es la función de distribución acumulativa para la distribución normal. Y la función $\Phi^{-1}(.)$ es la inversa. Luego, el intervalo de confianza BCPB está dado por la siguiente expresión.

$$[\hat{C}^*(P_{LEI} B), \hat{C}^*(P_{LES} B)] \tag{18}$$

RESULTADOS

Se llevaron a cabo diversas simulaciones a fin de obtener intervalos de confianza bootstrap de 95% para estimar los índices de capacidad de procesos con los métodos de Burr, Clements y Normal. En la figura 1, se presenta el histograma resultante de una muestra de tamaño 198 proveniente de una distribución Pareto Generalizada con parámetros $\hat{\alpha}=20,66$ y $k=0,44$

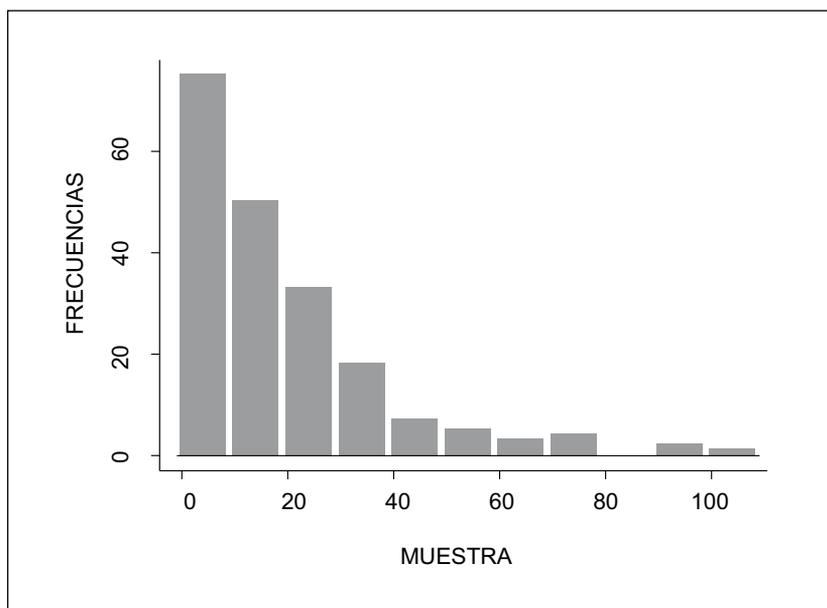


Figura 1. Histograma de la muestra de observaciones provenientes de DPG2(20,66; 0,44)

En la tabla 1, se presenta un resumen estadístico de las observaciones provenientes de la DPG2.

Tabla 1: Resumen estadístico de la muestra de observaciones provenientes de la DPG2(20,66;0,44).

Min	Q ₁	Mediana	Media	Q ₃	Max	Desviación	Asim	Kurt	$\hat{\alpha}$	\hat{k}
1,00	5,25	16,00	19,92	26,0	108,0	19,2314	1,89	4,47	20,66	0,44
Q _{0,00135}	Q _{0,99865}	$\alpha_3=1,87$	α_4							
1,0	105,34	1,87	7,33							

Se simulan 1000 muestras Bootstrap, sin reemplazamiento, obtenidas de la DPG(20,66;0,44) para determinar luego los valores estimados de los índices de capacidad que se tomarán como referencia. Se consideran tres diferentes valores targets para un C_{pu} los cuales son {0,5; 1,0 y 1,5}. Se define L como el valor mínimo de la muestra simulada. Con cada valor targets se obtiene el valor de U y los índices respectivos, los cuales se presentan en la tabla 2.

Tabla 2: Valores para diferentes C_{pu} , U y los índices de capacidad estimados

$C_{pu} = 0,50$	$C_{pu} = 1,0$	$C_{pu} = 1,5$
$U=60,67$	$U=105,34$	$U=150,01$
$\widehat{BC}_{pk=0,4885}$	$\widehat{BC}_{pk=0,9192}$	$\widehat{BC}_{pk=1,3499}$
$\widehat{CC}_{pk} = 0,4669$	$\widehat{CC}_{pk} = 0,8949$	$\widehat{CC}_{pk} = 1,3229$
$\widehat{C}_{pk} = 0,3280$	$\widehat{C}_{pk} = 0,3280$	$\widehat{C}_{pk} = 0,3280$

Seguidamente, se establecen tres métodos bootstrap para construir intervalos de confianza (Percentil, Estándar y el Percentil Sesgo-correcto) con un nivel de confianza de 95%.

Se simulan nuevamente 500 muestras de cada intervalo bootstrap obtenidos por las fórmulas (16)-(18) y se mide, en cada caso, el porcentaje de cubrimiento, la longitud promedio y desviación estándar de dichas longitudes promedios.

Los intervalos Bootstrap generados para los valores estimados de los \widehat{BC}_{pk} , \widehat{CC}_{pk} y \widehat{C}_{pk} en cada caso, se presentan en las tablas 3-5.

Es de hacer notar, que los intervalos obtenidos con el método tradicional utiliza para la estimación del índice C_{pk} a la media muestral como estimador, por lo que el C_{pk} no va a variar en resultado aun cuando varíen los valores de C_{pu} y U respectivamente.

A continuación se presentan las figuras 2, 3 y 4 correspondientes al cubrimiento de cada intervalo obtenido en la tabla 3.

Tabla 3: Proporción del cubrimiento para el 95% de límites de confianza bootstrap para el \widehat{BC}_{pk} bajo la DPG

$\widehat{BC}_{pk=0,4885}$	Cubrimiento	Ancho promedio	Desviación promedio
Percentil: [0,4092;0,5992]	0,97	0,505	0,0378
Estándar: [0,4005;0,5876]	0,94	0,493	0,0366
Percentil Corregido: [0,4068;0,5927]	0,93	0,499	0,0367
$\widehat{BC}_{pk=0,9192}$			
Percentil: [0,7809;1,1308]	0,99	0,9584	0,0691
Estándar: [0,7561;1,1062]	0,97	0,9328	0,0691
Percentil Corregido: [0,7604;1,1167]	0,97	0,9409	0,0715
$\widehat{BC}_{pk=1,3499}$			
Percentil: [1,1453;1,6680]	0,94	1,4033	0,1035
Estándar: [1,1204;1,6184]	0,97	1,3659	0,0978
Percentil Corregido: [1,1347;1,6278]	0,97	1,3837	0,0972

Tabla 4: Proporción del cubrimiento para el 95% de límites de confianza bootstrap para el \widehat{CC}_{pk} bajo la DPG

$\widehat{CC}_{pk=0,4669}$	Cubrimiento	Ancho promedio	Desviación promedio
Percentil: [0,3899;0,5725]	0,95	0,4794	0,0364
Estándar: [0,3824;0,5633]	0,98	0,4722	0,0352
Percentil Corregido: [0,3802;0,5574]	0,93	0,4643	0,0352
$\widehat{CC}_{pk=0,8949}$			
Percentil: [0,7542;1,0981]	0,92	0,9255	0,0673
Estándar: [0,7375;1,0819]	0,92	0,9107	0,0674
Percentil Corregido: [0,7419;1,0794]	0,97	0,9112	0,0665
$\widehat{CC}_{pk=1,3229}$			
Percentil: [1,1255;1,6274]	0,91	1,3788	0,0986
Estándar: [1,0934;1,6019]	0,96	1,3463	0,1005
Percentil Corregido: [1,1036;1,5871]	0,94	1,3465	0,0958

Tabla 5: Proporción del cubrimiento para el 95% de límites de confianza bootstrap para el \widehat{C}_{pk} bajo la DPG

$\widehat{CC}_{pk=0,3280}$	Cubrimiento	Ancho promedio	Desviación promedio
Percentil: [0,294; 0,3768]	0,87	0,3348	0,0165
Estándar: [0,29; 0,3739]	0,85	0,3321	0,0165
Percentil Corregido: [0,2884; 0,3689]	0,83	0,3287	0,0158

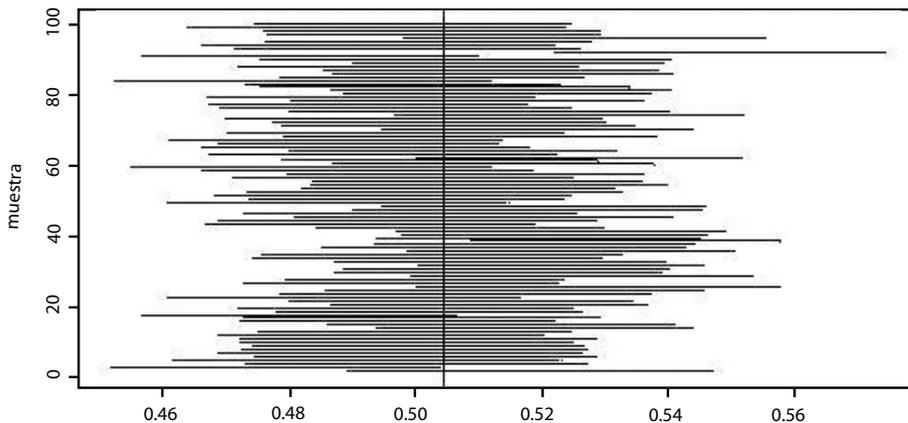


Figura 2. Cubrimiento de los intervalos Percentil Bootstrap

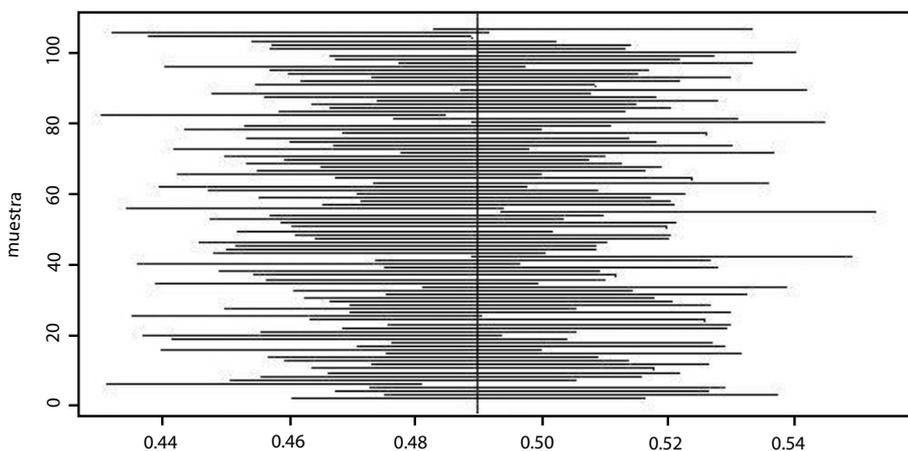


Figura 3. Cubrimiento de los intervalos Estándar Bootstrap

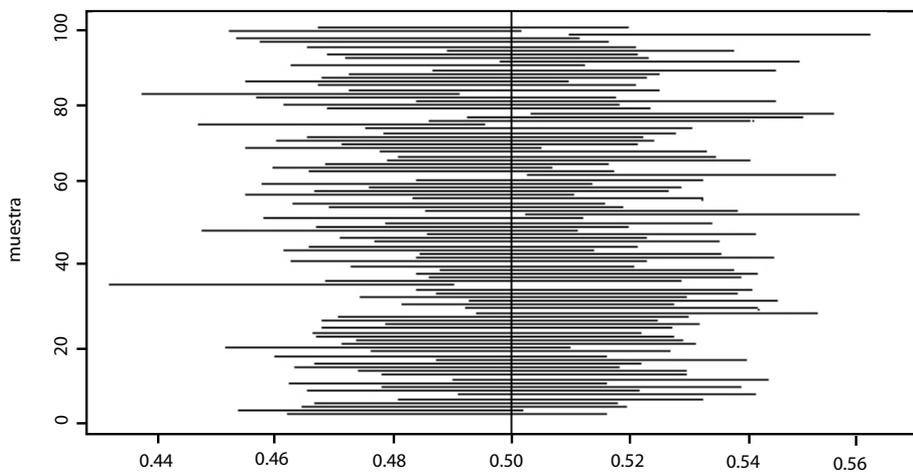


Figura 4. Cubrimiento de los intervalos Percentil Sesgo-Corregido Bootstrap.

DISCUSIÓN DE RESULTADOS

Para las muestras simuladas con la distribución PG2(20,66; 0,44) se lograron estimar los intervalos bootstrap dados por las fórmulas (16)-(18). El procedimiento se repitió de manera sistemática tomando diferentes valores targets del C_{pu} , {0,5; 1,0; 1,5}.

Para $C_{pu}=0,5$ se lograron los resultados BC_{pk} 0,4885 ; CC_{pk} 0,4669 y $\hat{C}_{pk}=0,3280$. De acuerdo a lo observado en las tablas (3)-(5), el PB para el BC_{pk} presenta mayor porcentaje de cubrimiento, mientras que \hat{C}_{pk} tiene menor porcentaje, así como el ancho promedio mayor entre los tres intervalos lo tiene el índice estimado BC_{pk} ; en cuanto a la desviación promedio, es mayor en los índices normales que en el índice normal. En el intervalo EB, el porcentaje de cubrimiento resulta mayor en el CC_{pk} que en los otros dos índices, así como su ancho promedio; las desviaciones son relativamente iguales entre los índices no normales, pero significativamente alta en comparación con el índice \hat{C}_{pk} estimado. Para el intervalo PSCB se mantienen los porcentajes de cubrimientos relativamente iguales entre los índices no normales, pero aún más alto en comparación con el índice \hat{C}_{pk} estimado; algo similar ocurre en la desviación promedio, la cual es más alta en los índices no normales. En general el PB es mejor estimador en promedio según los métodos de percentiles de Burr y de Clements, es más amplio que los intervalos EB y PSCB.

Por otro lado, para $C_{pu}=1,0$ se lograron los resultados $BC_{pk}=0,9192$; $CC_{pk}=0,8949$ y $\hat{C}_{pk}=0,3280$. El intervalo EB continúa teniendo mayor porcentaje en cubrimiento el índice estimado BC_{pk} al igual que el ancho promedio. Las desviaciones promedios son similares entre los estimados no normales mientras que el estimado bajo suposición normal se mantiene igual en todo los casos.

Finalmente, para $C_{pu}=1,5$ se lograron los resultados $BC_{pk}=1,3499$; $CC_{pk}=1,3220$ y $\hat{C}_{pk}=0,3280$, observándose que sigue siendo el índice estimado BC_{pk} el de mayor porcentaje en cubrimiento en los tres tipos de intervalos y el de mayor ancho promedio. Las desviaciones se mantienen muy similares entre los índices no normales, pero mayor en comparación con el índice normal. Para los diferentes valores targets fijados para C_{pu} (0,50; 1,0; 1,5), el intervalo PSCB es más confiable, en promedio, según los métodos de Percentiles de Burr y de Clements.

CONCLUSIONES

De acuerdo a las simulaciones realizadas, se puede generalizar que de los tres tipos de método utilizados para evaluar los índices de capacidad de procesos, el método de percentil de Burr arrojó mejores resultados que el método de percentil de Clements en lo que al porcentaje de cubrimiento y ancho promedio se refiere. Las dispersiones entre los índices no normales se mantienen significativamente iguales, por el contrario existe una variabilidad notable entre los índices estimados no normales y normales. Otro detalle que resalta en las estimaciones del índice C_{pk} normal es que éste mantiene constante para los diferentes valores que toma el C_{pu} y U, esto posiblemente porque la media de la muestra es mayor que la mediana y al evaluarlos en la fórmula (3) siempre se va a tomar el valor más pequeño. Otra razón es la fuerte asimetría de la muestra considerada (ver figura 1) la cual genera resultados erróneos al evaluar el índice suponiendo que la distribución es normal.

REFERENCIAS

- Burr, IW. (1942).** Cumulative frequency distribution. *Ann Math Stat.* 13:215-232.
- Burr, IW. (1973).** Parameters for a general system of distributions to match a grid of α_3 y α_4 . *Commun Stat.* 2;1-21.
- Chou, Y., Polanski, A. & Mason, R. (1998).** Transforming Non-normal data to normality in statistical Process Control. *Journal of Quality Technology.* 30;133-141.

- Clements, J.A. (1989).** Process Capability Calculations for Non-Normal Distribution, *Quality Progress*, 24 (9); 95-100.
- Davinson, A. (1984).** Modeling excesses over high Threshold, with Applications. In: *Statistical Extremes and Applications*, ed. J. Tiago de Oliveira, 461-482. Reidel, Dordrecht, the Netherlands.
- Efron, B. & Tibshirani, R. (1986).** Bootstrap Methods for Standard Errors Confidence Intervals, and Other Measures of Statistical Accuracy. *Statistical Science*, 1; 54-77.
- Franklin, L. & Wasserman, G. (1992).** Bootstrap Lower Confidence Limits for Capability Indices. *Journal of Quality Technology*. 24, 196-210.
- Gruska, G.; Mirkhani, K. & Lamberson, L. (1989).** Non-normal Data Analysis.(St. Clair Shores, MI: Applied Computer Solutions, Inc.).
- Han, C. (2006).** Effect of Testing Normality on Estimating Process Capability Indices. *Quality Engineering*, 18; 391-395.
- Hosking, J. & Wallis, J. (1987).** Parameter and Quantile Estimation for the Generalized Pareto Distribution. *Technometrics*, 29(3); 339-349.
- Kane, V. (1986).** Process Capability Indices. *Journal of Quality Technology*, **18**; 41-52.
- Kotz, S. & Johnson N. (1993).** Process Capability Indices. Chapman and Hall. London.
- Liu, P. & Chen, F. (2006).** Process Capability Analysis on Non-normal Process Data Using the Burr XII Distribution. *The International Journal of Advanced Manufacturing Technology*, 27; 975-984.
- Pickands, J. (1975).** Statistical Inference Using Extreme Order Statistics. *Ann. Statist.* 3; 119-131.
- Pal, S. (2005).** Evaluation of non-normal process Capability Indices using Generalized Lambda Distribution. *Quality Engineering*. 17;77-85.
- Polanski, A. (1998).** A Smooth Nonparametric Approach to Process Capability. *Quality and Reliability Engineering International*. 14; 43-48.
- Saksena, S. & Johnson, A. (1984).** *Estimators Best Umbiased for the Parameters of a two-parameter Pareto Distribution*. *Biometrika*, 31; 77-83.
- Van Montfort, M. & Witter, J. (1985).** *Testing exponentially against Generalized Pareto Distributions*. *J. Hydrol.* 78; 305-315._