

Estudio comparativo de índices de capacidad de proceso con variables distribuidas no normales

Recibido: 14/06/10 Aceptado: 02/07/10

⁽¹⁾Martha Valdiviezo
⁽²⁾José Fermín

RESUMEN

En este trabajo se evalúa el desempeño de índices de capacidad de procesos (ICP), en situaciones no normales utilizando los métodos de percentiles de Clements (CC_p, CC_{pk}) y de Burr (BC_p, BC_{pk}). Aunque los ICP se utilizan en la industria, no existe suficiente literatura para determinar su precisión tomando en cuenta desviaciones moderadas y severas de normalidad. Para estudiar estas desviaciones, se lleva a cabo una comparación de ambos métodos considerando datos simulados con las distribuciones Weibull, Log normal, Beta y Pareto Generalizada (PG). Al calcular los ICP, el CC_{pk} arroja desviaciones más pequeñas con la distribución Weibull, mientras que con el BC_{pk} , las desviaciones son mayores. Por otro lado, con cada distribución, en promedio el BC_p genera menor desviación con respecto al CC_p . Finalmente se presenta un caso real considerando la distribución PG.

Palabras clave: Índices de capacidad de procesos no normales, distribuciones Burr y Pareto generalizada, percentiles de Clements y Burr

COMPARATIVE STUDY OF PROCESS CAPABILITY INDICES VARIABLES DISTRIBUTED WITH NO NORMAL

ABSTRACT

This paper evaluates the performance of process capability indices, PCI in non-normal situations percentiles using the methods of Clements (CC_p, CC_{pk}) and Burr (BC_p, BC_{pk}). Although the PCI is used in industry, there is insufficient literature to determine their accuracy by taking into account moderate and severe deviations of the normality. To study these deviations, it performs a comparison of both methods considering simulated data distributions with the Weibull, Lognormal, Beta and Generalized Pareto, GP. In calculating the PCI the CC_{pk} generates smaller deviations with the Weibull distribution, whereas that the BC_{pk} deviations are higher. On the other side with each distribution the BC_p generates lower average deviation from the CC_p . Finally a real case considering the distribution GP.

Keywords: Process Capability Indices. Burr and Generalized pareto distributions, Burr & Clements percentiles.

INTRODUCCIÓN

Los estudios de capacidad de procesos (ECP), se usan para el monitoreo de la capacidad de un proceso. Esto se traduce en la recolección de datos provenientes del proceso, a fin de tener una justa visión de su desempeño. El desempeño de dicho proceso puede evaluarse mediante el uso de índices de capacidad de procesos (ICP), los cuales proporcionan medidas numéricas de la capacidad y mejoramiento de dicho proceso.

Kane (1986), dio las primeras explicaciones e interpretaciones de estos índices. Una definición de ICP es la siguiente; para una variable del proceso X, independiente e idénticamente distribuida, con una distribución normal con media μ y desviación estándar s . Entonces los índices se definen de la siguiente manera:

$$C_p = \frac{U - L}{6\sigma} \quad (1)$$

$$C_{pl} = \frac{\mu - L}{3\sigma} \quad (2)$$

$$C_{pu} = \frac{U - \mu}{3\sigma} \quad (3)$$

$$C_{pk} = \min [C_{pu}, C_{pl}] \quad (4)$$

donde L y U representan los límites de especificación inferior y superior, respectivamente, del proceso; el valor target (T) está en el punto medio de los límites de especificaciones. Cuando el proceso está en control estadístico, estos índices se estiman sustituyendo μ y s , respectivamente, por \bar{x} y s .

Aparte de que el proceso en estudio debe estar estable estadísticamente hablando, otro importante supuesto de la capacidad de un proceso y de estos índices es que su interpretación usual se basa en que la salida del proceso sigue una distribución normal; pero si la distribución fundamental es no-normal, entonces los enunciados acerca de la fracción no-conforme esperada atribuida a un valor particular de C_p o C_{pk} pueden generar resultados erróneos.

1 M.Sc en Matemática. Profesora de la Escuela de Administración de la UDO. Núcleo de Sucre. Cumaná-Venezuela. E-mail: mvaldivi70@cantv.net
 2 Ph.D. in Statistics, Kansas University. EEUU. Profesor Jubilado del Instituto Universitario de Tecnología Cumaná. Venezuela. E-mail: simon.fermin@gmail.com

En la industria, existen muchos procesos de manufactura donde la variabilidad del proceso no se puede describir por una distribución normal. En este caso no se deberían usar las mismas fórmulas para calcular los índices de capacidad de procesos, porque los resultados darían resultados engañosos. Somerville y Montgomery (1997), reportaron que el uso de índices de capacidad de procesos normales en vez de no normales, generan resultados no confiables para este tipo de procesos, sugirieron que no se debe utilizar las mismas fórmulas para el cálculo de los índices de capacidad de procesos tradicionales debido a que estos, producirían resultados engañosos acerca del desempeño del proceso en cuestión.

Un enfoque para tratar con datos no normales es transformarlos utilizando alguna función matemática, de tal manera que los datos resultantes sean normales. Según Chou *et al.* (1998), Polanski *et al.* (1998) y Pal (2005), la idea es ajustarlos mediante una distribución no normal tales como las distribuciones Weibull, Gamma o Beta para luego estimar los cuantiles, los cuales se pueden usar para calcular los índices de capacidad de procesos.

Por otro lado, para estimar ICP con datos distribuidos no normales se puede utilizar el método de percentil de Clements (1989). Este método se utiliza para calcular los índices C_p y C_{pk} mediante una familia de curvas de Pearson. Consiste en estimarlos mediante una simple modificación de los ICP normales, utilizando los valores de los puntos percentiles de distribuciones no normales y utilizando la mediana como medida central y no la media. Por otro lado, Liu y Chen (2006), introducen una modificación del método de Clements, sustituyendo los percentiles de las familias de distribuciones de Pearson por una distribución Burr XII apropiada.

En este trabajo, se realiza un estudio comparativo del desempeño de los ICP no normales utilizando los métodos de Percentil de Clements y Percentil de Burr, tomando en cuenta las distribuciones, Weibull, Log-normal, Beta y la Pareto Generalizada (PG), todas con dos parámetros. Se toma en cuenta primeramente un estudio de simulación con las distribuciones antes mencionadas para observar el desempeño de los ICP; seguidamente se considera un conjunto de datos reales obtenidos de un estudio hidrológico cuya distribución resulta ser PG, se estiman los parámetros de escala y forma usando el método de máxima verosimilitud, se determinan los ICP y finalmente se comparan los resultados.

MÉTODOS PARA ESTIMAR LOS ÍNDICES DE CAPACIDAD PARA PROCESOS DISTRIBUIDOS NO-NORMALES

En esta sección se revisan brevemente dos métodos que se utilizarán para estimar los ICP con datos distribuidos no normales.

Método de Percentil de Clements

Este método es muy utilizado en la industria, tal como ya se mencionó anteriormente, 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 (1), el valor 6s 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 $U_p - L_p$.

$$CC_p = \frac{U - L}{U_p - L_p} \tag{5}$$

donde, U_p es el percentil 99,865 y L_p es el percentil 0,135 presentado por Gruska *et al.* (1989).

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

$$CC_{pl} = \frac{M - L}{M - L_p} \tag{6}$$

$$CC_{pu} = \frac{U - M}{U_p - M} \tag{7}$$

$$CC_{pk} = \text{Min}(CC_{pu}, CC_{pl}) \tag{8}$$

donde, M, es la mediana de la distribución del proceso, L y U son los límites de especificación inferior y superior, respectivamente; los valores L_p y U_p son los cuantiles 0,135% y 99,865% de la distribución utilizada.

Este método utiliza el estimador clásico de simetría y kurtosis los cuales están basados en el tercer y cuarto momento respectivamente, estos algunas veces pueden no ser muy confiables para muestras pequeñas. Tanto el tercer como el cuarto mo-

mento 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 Percentil 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.

$$f(x|c,k) = \begin{cases} c k x^{c-1} (1+x^c)^{-(k+1)} & \text{si } x \geq 0, c \geq 1, k \geq 1 \\ 0 & \text{si } x < 0 \end{cases} \quad (9)$$

donde, c y k representan los coeficientes de asimetría y la kurtosis de la distribución Burr. Por otro lado, la función de distribución acumulada y la función de transformación recíproca de la distribución Burr XII, están dadas respectivamente, por:

$$F(x|c,k) = \begin{cases} 1 - (1+x^c)^{-k} & \text{si } x \geq 0, c \geq 1, k \geq 1 \\ 0 & \text{si } x < 0 \end{cases} \quad (10)$$

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

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. Se pueden emplear los primeros cuatro momentos de una distribución empírica. El tercer y cuarto momento de una distribución empírica cualquiera (s₃ y s₄), están dados por las siguientes expresiones matemáticas, respectivamente:

$$s_3 = \frac{n}{(n-1)(n-2)} \sum_i \left(\frac{x_i - \bar{x}}{s} \right)^3 \quad (12)$$

$$s_4 = \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum_i \left(\frac{x_i - \bar{x}}{s} \right)^4 - \frac{3(n+1)^2}{(n-2)(n-3)} \quad (13)$$

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

Los valores tabulados de asimetría y kurtosis representan una transformación estandarizada entre una variable Burr, U, y otra variable aleatoria, X.

Liu y Chen (2006), introdujeron una modificación del método de Clements el cual utiliza percentiles en una familia de curvas de Pearson. En este método se sustituyen estos puntos percentiles 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 (s₃) y la kurtosis (s₄) 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 Z_{0,00135}, Z_{0,5} y Z_{0,99865}, respectivamente.

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; se calculan de la siguiente manera.

$$L_p = X_{0,00135} = \bar{x} + s_x Z_{0,00135} \quad (14)$$

$$M = X_{0,50} = \bar{x} + s_x Z_{0,50} \quad (15)$$

$$U_p = X_{0,99865} = \bar{x} + s_x Z_{0,99865} \quad (16)$$

Paso 5: Finalmente se estiman los índices de capacidad de procesos mediante las expresiones matemáticas dadas en (5)-(8).

DISTRIBUCIONES UTILIZADAS PARA EL ESTUDIO DE SIMULACIÓN

Distribución Pareto Generalizada con 2 parámetros

La distribución Pareto generalizada con 2 parámetros, DP(2), fue presentada por Pickands (1975), y desde entonces su aplicabilidad ha sido uti-

lizada en muchas áreas, sobre todo en el campo de la economía. Hosking y Wallis (1987), presentan una definición de la DPG2, tomando en cuenta una variable aleatoria X, definida de la manera siguiente: $X = \frac{\hat{a}(1 - e^{-kY})}{k}$; donde Y es una variable aleatoria con distribución exponencial estándar, a y k son los parámetros.

La distribución de X se dice que es una DPG2, si la función de distribución está definida de la siguiente manera:

$$F(x) = \begin{cases} 1 - \left(1 - \frac{kx}{\hat{a}}\right)^{1/k} & \text{si } k \neq 0 \\ 1 - \exp\left(-\frac{x}{\hat{a}}\right) & \text{si } k = 0 \end{cases} \quad (17)$$

Así mismo, la función de densidad de probabilidad está dada de la siguiente manera:

$$f(x) = \begin{cases} \hat{a}^{-1} \left(1 - \frac{kx}{\hat{a}}\right)^{1/k - 1} & \text{si } k \neq 0 \\ \hat{a}^{-1} \exp\left(-\frac{x}{\hat{a}}\right) & \text{si } k = 0 \end{cases} \quad (18)$$

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 $\alpha > 0$, $0 \leq x < \alpha/k$. Cuando $k = 0$, se obtiene la distribución exponencial con media a y cuando $k = 1$ se tiene la distribución uniforme sobre [0,a]. Cuando $k < 0$, esta distribución se denomina simplemente Pareto. Para $k > 1/2$, la distribución PG presenta finitos puntos finales para $f(x) > 0$.

La función de verosimilitud está dada de la siguiente manera.

$$\text{Log } L(x; a, k) = \begin{cases} -n \log \alpha - \left(1 - \frac{1}{k\alpha}\right) \sum_{i=1}^n \log\left(1 - \frac{kx_i}{\alpha}\right) & \text{para } k \neq 0 \\ n \log \alpha - \left(1 - \frac{1}{k\alpha}\right) \sum_{i=1}^n \frac{x_i}{\alpha} & \text{para } k = 0 \end{cases} \quad (19)$$

El rango de a es positivo para $k \leq 0$ y $a > kx_{(n)}$, para $k > 0$. Los parámetros a y k pueden o no existir.

Distribución Weibull (We)

Se dice que una variable aleatoria X, tiene una distribución de Weibull (We), si su función de densidad de probabilidad con parámetros de forma a y de escala b está dada de la siguiente forma.

$$f(x, \alpha, \beta) = \begin{cases} \frac{\alpha}{\beta^\alpha} (x^{\alpha-1}) \exp\left[-\left(\frac{x}{\beta}\right)^\alpha\right] & \text{para } x \geq 0; \alpha, \beta > 0 \\ 0 & \text{para cualquier otro valor} \end{cases} \quad (20)$$

Distribución Log-normal (Ln)

Sea X una variable aleatoria distribuida normalmente con media m y varianza s^2 . La función de densidad de probabilidad para una variable de la forma, $Y = \exp(X)$, está dada de la siguiente manera.

$$f_y(y, \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma y}} \exp\left\{-\frac{1}{2} \left[\frac{\ln(y) - \mu}{\sigma}\right]^2\right\} \quad \text{para } y > 0 \quad (21)$$

con, $-\infty < \mu < \infty$ y $x \geq 0$.

Distribución Beta (Be)

Sea X una variable aleatoria, se dice que X tiene una distribución Beta, (Be), si la función de densidad de probabilidad está dada por:

$$f(x, \alpha, \beta) = \begin{cases} \frac{\Gamma(\alpha + \beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1} (1-x)^{\beta-1} & \text{para } 0 < x < 1; \alpha, \beta > 0 \\ 0 & \text{para cualquier otro valor} \end{cases} \quad (22)$$

ESTUDIO DE SIMULACIÓN

Se considera en este estudio varias distribuciones no normales para simular conjuntos de datos, ellas son: Beta, Log-normal, Weibull y Pareto Generalizada. Estas distribuciones se utilizan para investigar el efecto que tiene la no normalidad en los datos al estimar los índices de capacidad de procesos. Además, dichas distribuciones son conocidas (de acuerdo a los valores de los parámetros escogidos), porque representan desviaciones que van desde severas hasta moderadas de la normalidad. Por ejemplo para muestras simuladas de tamaño 1000, una distribución Ln(0,5;0,2) genera desviación leve, una Ga(4,0 ; 0,5), genera desviación mo-

derada y una distribución We (1,0; 1,2), genera una desviación severa de la normalidad (Ver Figura 1).

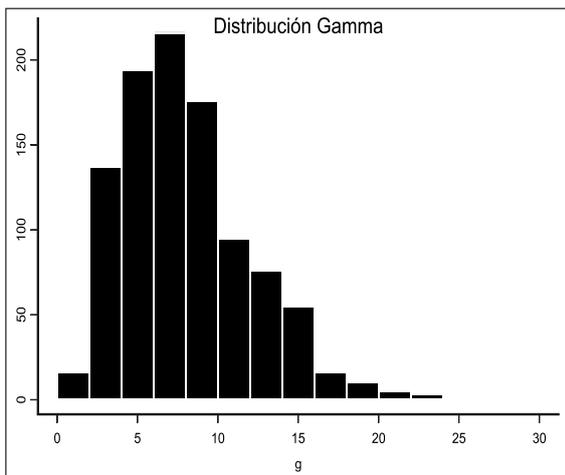
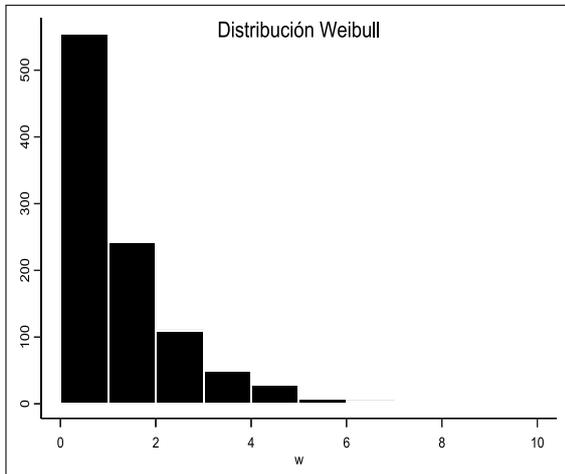
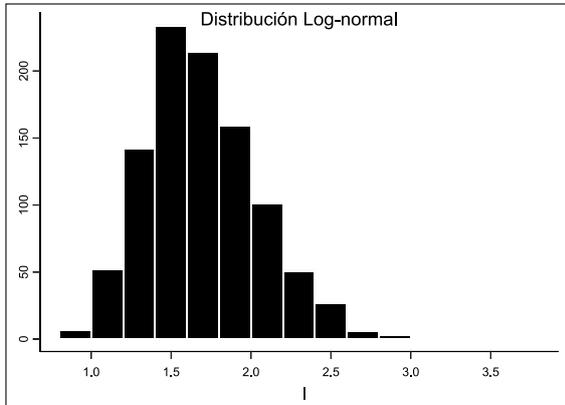


Figura 1. Distribuciones Ln (0,5;0,2); Ga (4,0;0,5) y We (1,0;1,2) para una muestra de tamaño 1000.

SIMULACIONES CON DIFERENTES DISTRIBUCIONES

Para este estudio, se consideran corridas de 30 muestras cada una de tamaño 100, se consideran los valores teóricos para diferentes valores de C_{pu} : {1,0; 1,5; 2,0}. Se toman las distribuciones We(1,0;1,2); Ln(0;1); Be(4,4;13,3) y PG2(0,1;0,2). Se calcula el valor del límite de especificación superior U el cual será el mismo para cada C_{pu} , mediante la siguiente expresión.

$$U = C_{pu} (X_{0,99865} - X_{0,5}) + X_{0,5}$$

donde, $X_{0,99865}$ y $X_{0,5}$ son los puntos percentiles en cada distribución. Además, se calculan los valores de las medias, desviaciones estándares, coeficientes de asimetrías y kurtosis en cada conjunto de datos. Utilizando los métodos de Percentil de Clements y Burr se estiman los valores de los índices de capacidad de procesos dados por las fórmulas (5) hasta la (8). Se estiman los valores promedios de U, medias y desviaciones muestrales, se determinan los percentiles estandarizados de la distribución Burr y los percentiles correspondientes al método de Clements. Finalmente se determinan los valores promedios de los índices de Clements, CC_{pu} y $CC_{p\sigma}$, similarmente los índices de acuerdo al método de Burr, BC_p y $BC_{p\sigma}$.

El procedimiento se resume de la siguiente manera:

Se escoge una distribución con sus respectivos parámetros (Weibull, Log-normal, Beta y PG2); se simulan 30 muestras de tamaño 100. Se determinan los valores de las medias, desviaciones estándar, coeficientes de asimetría y kurtosis; con los valores antes calculados se determinan los percentiles estandarizados de las curvas de Pearson, $X_{0,00135}$; $X_{0,5}$ y $X_{0,99865}$ y los percentiles de la distribución Burr $B_{0,00135}$; $B_{0,5}$ y $B_{0,99865}$ respectivamente.

De acuerdo a esto, se calculan los valores de L_p , U_p de acuerdo a las fórmulas (14)-(16) en cada método. Finalmente se estiman los índices de capacidad con las fórmulas (5) hasta la (8).

Los resultados obtenidos para cada distribución se presentan en las tablas 1 hasta la 4 respectivamente. De manera similar, se presentan en las figuras 2 hasta la 5 las gráficas de cajas para cada distribución en estudio.

Tabla 1. Resultados promedios para la Distribución We (1,0; 1,2).

C _{pu}	U	Burr		Clements	
		BC _p		CC _p	
		Media	Desviación	Media	Desviación
1,0	5,7061	0,8876	0,0777	0,9058	0,08798
1,5	8,1309	1,2723	0,1183	1,2981	0,1290
2,0	10,556	1,6570	0,1597	1,6903	0,1711

C _{pu}	U	BC _{pk}		CC _{pk}	
		Media	Desviación	Media	Desviación
		1,0	5,7061	0,8163	0,1311
1,5	8,1309	1,1219	0,3040	0,9718	0,2459
2,0	10,556	1,3478	0,4776	1,0057	0,3261

Tabla 3. Resultados promedios para la Distribución Log-normal (0;1).

C _{pu}	U	Burr		Clements	
		BC _p		CC _p	
		Media	Desviación	Media	Desviación
1,0	5,7061	0,9212	0,08796	0,9489	0,0862
1,5	8,1309	1,3282	0,13094	1,3679	0,1261
2,0	10,556	1,7351	0,1744	1,7244	0,2801

C _{pu}	U	BC _{pk}		CC _{pk}	
		Media	Desviación	Media	Desviación
		1,0	5,7061	0,8672	0,0827
1,5	8,1309	1,2866	0,1217	1,1838	0,1996
2,0	10,556	1,6712	0,2134	1,3643	0,3944

Tabla 2. Resultados promedios para la Distribución Be (4,4; 13,3).

C _{pu}	U	Burr		Clements	
		BC _p		CC _p	
		Media	Desviación	Media	Desviación
1,0	5,7061	0,9137	0,1209	0,9192	0,1215
1,5	8,1309	1,2029	0,1579	1,2293	0,1714
2,0	10,556	1,3721	0,3056	1,5394	0,222

C _{pu}	U	BC _{pk}		CC _{pk}	
		Media	Desviación	Media	Desviación
		1,0	5,7061	0,7987	0,1209
1,5	8,1309	1,0459	0,2949	0,9278	0,3117
2,0	10,556	1,1585	0,4347	0,9865	0,4333

Tabla 4. Resultados promedios para la Distribución PG2 (0,1; 0,2).

C _{pu}	U	Burr		Clements	
		BC _p		CC _p	
		Media	Desviación	Media	Desviación
1,0	5,7061	0,8031	0,0853	0,8149	0,0993
1,5	8,1309	1,1898	0,12695	1,2070	0,1469
2,0	10,556	1,5764	0,1689	1,5366	0,3014

C _{pu}	U	BC _{pk}		CC _{pk}	
		Media	Desviación	Media	Desviación
		1,0	5,7061	0,6735	0,1727
1,5	8,1309	0,7409	0,2930	0,7021	0,3177
2,0	10,556	0,7849	0,4029	0,7405	0,4214

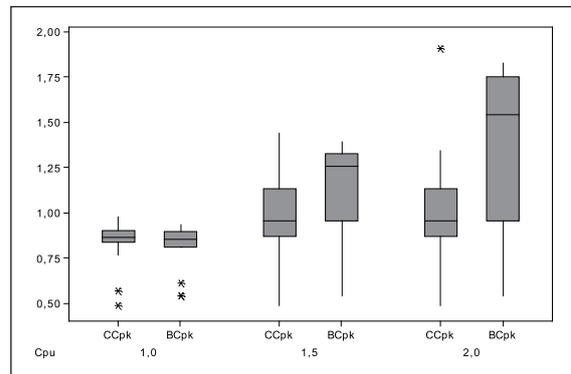
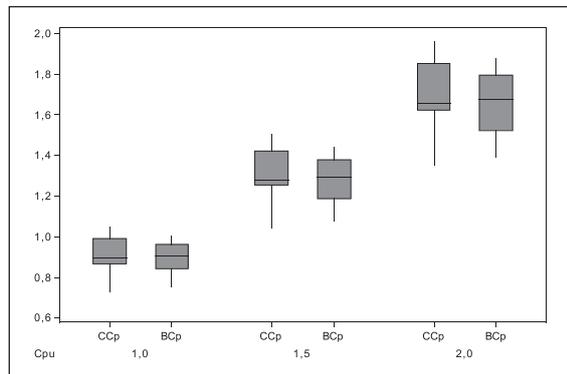


Figura 2. Diagrama de caja para los índices (CC_p, BC_p) y (CC_{pk}, BC_{pk}), respectivamente para una Distribución We (1,0;1,2).

ESTUDIO COMPARATIVO DE ÍNDICES DE CAPACIDAD DE PROCESOS CON VARIABLES DISTRIBUIDAS NO NORMALES

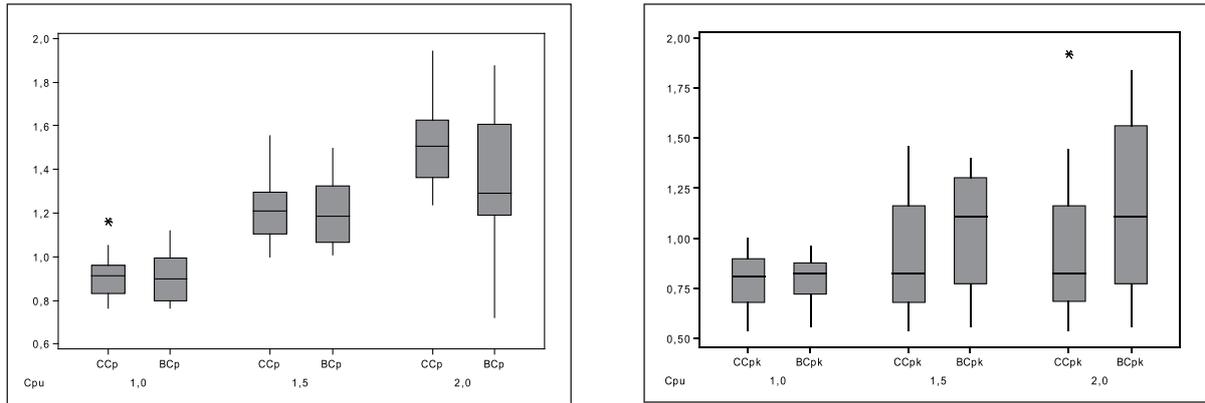


Figura 3. Diagrama de caja de los índices $(CC_p; BC_p)$ y $(CC_{pk}; BC_{pk})$, respectivamente, para la Distribución $Be(4,4;13,3)$.

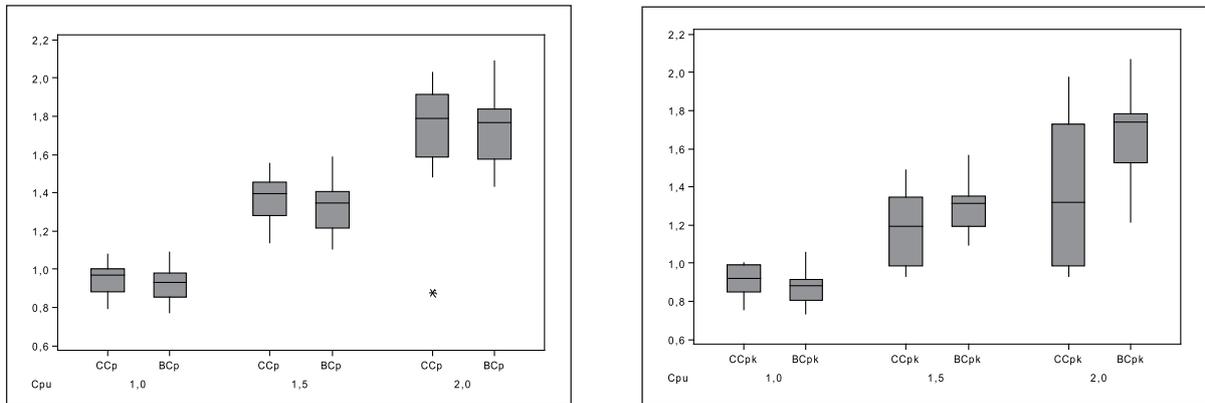


Figura 4. Diagrama de caja de los índices $(CC_p; BC_p)$ y $(CC_{pk}; BC_{pk})$, respectivamente, para la Distribución $Ln(0,1)$.

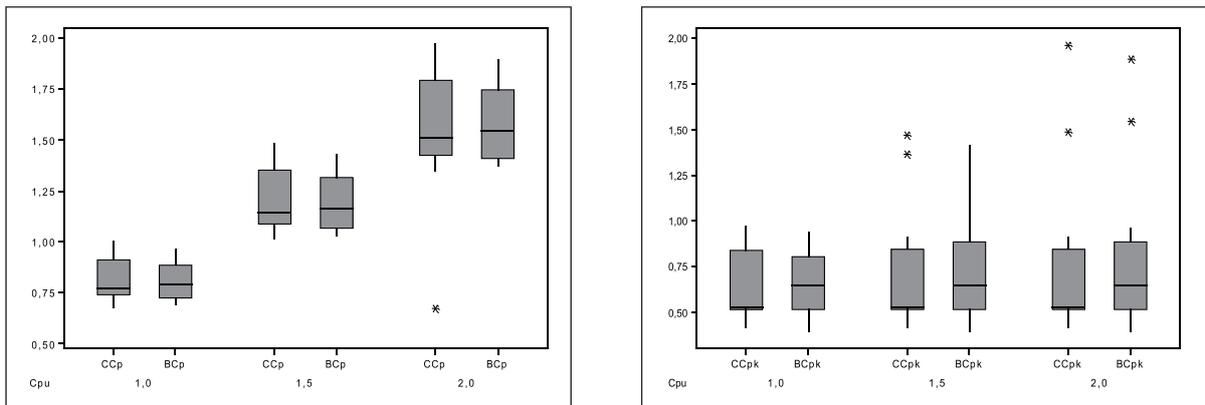


Figura 5. Diagrama de caja de los índices $(CC_p; BC_p)$ y $(CC_{pk}; BC_{pk})$, respectivamente, para la Distribución $PG2(0,1;0,2)$.

DISCUSIÓN

Para considerar el comportamiento de los índices obtenidos mediante los métodos de Clements (CC_p , CC_{pk}) y Burr (BC_p y BC_{pk}), respectivamente, se tomará en cuenta el efecto ocasionado sobre la desviación estándar obtenida en cada distribución; de acuerdo a los resultados presentados en las tablas 1-3 y comparando los índices de capacidad de procesos basados en los métodos de percentiles de Clements y Burr (CC_p vs BC_p) y (BC_{pk} vs CC_{pk}), se pueden observar varios resultados.

Respecto a la media y tomando en cuenta el ICP estimado mediante el método de percentiles de Clements CC_p y el de Burr, BC_p , para los diferentes C_{pu} la desviación tomando en cuenta las distribuciones Pareto Generalizada, Weibull, Beta y Log-normal se tiene el siguiente orden, $CC_{pPG2} < CC_{pWe} < CC_{pBe} < CC_{pLn}$ lo que indica que mientras más severa es la desviación, al suponer normalidad en los datos, los resultados obtenidos son menos confiables. Por otro lado, el ICP determinado con el método de percentil de Clements CC_p estima mejor los datos que el de Burr, es decir $CC_p > BC_p$. En cuanto a la precisión se tiene en general, las distribuciones con asimetría moderadas son menos precisa, el orden dado es el siguiente, $CC_{pLn} < CC_{pBe} < CC_{pWe} < CC_{pPG2}$. Similarmente, se tiene la relación entre los índices no normales $BC_{pk} > CC_{pk}$, es decir en promedio, la precisión de los ICP es mayor considerando el método de percentiles de Burr que en el Clements. Finalmente, la distribución PG2 (0,1; 0,2) arroja la desviación más pequeña considerando el método de Burr en contraparte con la distribución Log-normal que presenta desviaciones un poco mayor que las demás.

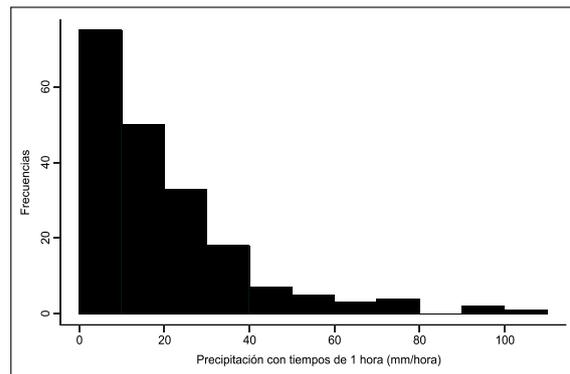
En los diagramas de cajas se muestran resultados como la mediana, primer y tercer cuartil, aparte de los outliers observados en cada distribución estudiada.

CASO REAL

En esta sección se ilustran ambos métodos en el ajuste de datos muestrales a una distribución desconocida. La distribución obtenida se utiliza para determinar los índices de capacidad de procesos. Se considera una serie de datos observacionales sobre intensidad máxima de precipitación diaria (mm/horas) medidas en la estación de M.A.R.N., estación Puerto la Cruz ubicada en el municipio Sotillo de la ciudad de Puerto la Cruz, Estado Anzoátegui Venezuela. Dicha información refleja las precipitaciones caídas en el estado con un tiempo máximo de duración de 1 hora, ocurridas durante los años

1961 hasta 2000. La serie contiene 198 observaciones. El objetivo es evaluar un proceso hidrológico ocurrido en el periodo de tiempo 1961-2000. La aleatoriedad de los datos se verifica para asegurar la estabilidad del proceso. Los datos muestrales se ordenan de manera ascendente. La distribución de frecuencia y el resumen se presentan en la Tabla 5.

Por otro lado, en la Figura 6 se representa el histograma de la serie de datos considerada.



Fuente: M.A.R.N., estación Puerto la Cruz. Municipio Sotillo.

Figura 6: Histograma de la serie de datos con tiempo de una hora máxima de precipitación.

Para verificar la bondad de ajuste de la distribución en estudio que resultó ser una DPG2, se siguen los siguientes pasos.

Paso 1: Los valores de los parámetros a y k como son desconocidos, estos se estimaron utilizando el método máxima verosimilitud usando S-plus, dichos resultados fueron los siguientes, $\hat{a}=20,655$ y $\hat{k}=0,037$.

Paso 2: Se llevó a cabo los test de Cramer Von Mises, W^2 , y Anderson Darling, A^2 , cuyos resultados fueron, $W^2=1,3676$ ($p < 0,005$) y $A^2=8,664$ ($p < 0,005$).

De la curva de distribución Pareto ajustada, se estiman los valores de los cuantiles 0,00135; 0,5 y 0,99865, respectivamente. Dichos valores son: $L_p=1,00$; $M=16,00$ y $U_p=105,341$. Además se tienen otros resultados que se presentan en la Tabla 5. Los límites de especificaciones inferior y superior considerados para este estudio fueron tomados en $L=1$ y $U=107$, respectivamente.

Paso 3: Observando los valores de los p-value se puede inferir que dicha serie de datos es ajustada por una distribución Pareto con parámetros a y k.

Paso 4: Finalmente, se estiman los índices de capacidad generalizados y los usuales utilizando las

fórmulas (1) hasta la (8). Los resultados se muestran en la tabla 6.

Tabla 5. Resumen estadístico para una serie de datos obtenida de un proceso hidrológico con un tiempo de duración de una hora de lluvia.

Min	Q ₁	Med	\bar{X}	Q ₃	Max	s	Asim	Kurtosis
1,00	5,25	16,00	19,92	26,00	108,00	19,23141	1,798	4,50
L _p	U _p	a ₃	a ₄	L	U	\hat{a}	\hat{k}	
1,00	105,341	1,87	7,3	1,00	107,00	20,655	0,037	

Tabla 6. Valores de los índices obtenidos con los métodos de Percentil de Clements, de Burr y los tradicionales, respectivamente.

Índices de capacidad usando Percentil de Clements	Índices de capacidad usando Percentil de Burr	Índices de Capacidad de procesos normales
$C\hat{C}_p^* = 0,955$	$B\hat{C}_p^* = 0,997$	$\hat{C}_p = 0,919$
$C\hat{C}_{pk}^* = 0,911$	$C\hat{C}_{pk}^* = 0,935$	$\hat{C}_{pk} = 0,328$

De acuerdo a los resultados obtenidos para los índices respectivos, se puede observar que la serie de datos proporcionan índices de capacidad para un proceso hidrológico no capaz, ya que estos son menores a los valores estándares. Se puede observar que los índices estimados con el método de percentiles de Clements ($C\hat{C}_p^*$ y $C\hat{C}_{pk}^*$) y con el método de percentiles Burr ($B\hat{C}_p^*$ y $B\hat{C}_{pk}^*$), son mayores que los tradicionales (\hat{C}_p y \hat{C}_{pk}). Por otro lado, el método de Burr genera mejores resultados cuando la distribución considerada presenta una desviación severa en los datos, esto permite asegurar que la asimetría en los datos observacionales es influyente a la hora de determinar dichos índices. Se aprecia la diferencia entre considerar una distribución normal de los datos y tomar una distribución asimétrica como lo es la distribución PG2.

CONCLUSIONES

En este trabajo se evalúa el desempeño de los índices de capacidad de procesos en situaciones no normales, simulando un conjunto de datos con distribuciones con asimetrías moderadas y severas (We, Be, Ln, PG2), para luego ser comparados entre sí. Se observa en general que el método de Burr arroja resultados con menor dispersión en las distribuciones con asimetría severa de su normalidad en comparación con el método de percentiles

de Clements. En cuanto a precisión, los ICP estimados para las distribuciones con asimetrías moderadas son menos confiables que los que calculados para distribuciones con asimetrías severas. Por otro lado, para el conjunto de datos reales se logra ajustar la distribución PG2, tomando en cuenta dos test de bondad de ajuste para determinar cual distribución es la adecuada a los datos, se determinan sus dos parámetros (a, k) mediante el método de estimación por máxima verosimilitud y finalmente se calculan los ICP bajo los dos enfoques es decir, cuando no son normales y suponiendo que sí lo son (fórmulas (1) - (8)), observándose en la tabla 6, que los índices obtenidos con el método de percentiles de Clements ($C\hat{C}_p^*$ y $C\hat{C}_{pk}^*$), y los obtenidos con el método de Burr ($B\hat{C}_p^*$ y $B\hat{C}_{pk}^*$) arrojan mejores resultados, mientras que para los índices calculados con el método tradicional (\hat{C}_p y \hat{C}_{pk}) los resultados son más pequeños. Esto indica que a la hora de tomar la decisión de usar los ICP bajo el supuesto de normalidad para un proceso determinado se obtendrían resultados erróneos.

REFERENCIAS BIBLIOGRÁFICAS

- [1] Burr IW. (1942). "Cumulative frequency distribution". *Ann Math Stat.* 13:215-232.
- [2] Burr IW. (1973). "Parameters for a general system of distributions to match a grid of a₃ y a₄". *Commun Stat.* 2:1-21.
- [3] Baxter, M. (1980). "Minimum variance unbiased estimation of the parameter of the Pareto Distributions". *Biometrika*, 27,133-138.
- [4] Clements, J. (1989). "Process Capability Calculations for Non-Normal Distribution", *Quality Progress*, 24 (9), 95-100.
- [5] Chou, Y., Polanski, A. & Mason, R. (1998). "Transforming Non-normal data to normality in statistical Process Control". *Journal of Quality Technology*. 30:133-141.
- [6] Cook, W. & Mumme, D. (1981). "Estimation of Pareto parameters by numerical methods". In: *Statistical Distributions in scientific Work*, ed. C. Tillie et al. 5, 127-132.
- [7] Davinson, A. & Smith, R. (1990). Models for Exceedances Over High Thresholds. *J. Roy. Statis. Soc. B.* 52(3), 393-442.
- [8] Hosking, J. & Wallis, J. (1987). "Parameter and Quantile Estimation for the Generalized Pareto Distribution". *Technometrics*, Vol. 29, N.º 3. 339-349.

- [9] Kane, V. (1986). "Process capability indices". *Journal of Quality Technology*. 18. 41-52.
- [10] Kotz, S. and Johnson N. (1993). "Process capability indices". London: Chapman and Hall.
- [11] 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.
- [12] Pal, S. (2005). "Evaluation of non-normal process capability indices using generalized lambda distribution". *Quality Engineering*. 17:77-85. 77-85.
- [13] Pickands, J. (1975). "Statistical inference using extreme order statistics". *Ann. Statist.* 3. 119-131.
- [14] Polanski, A. (1998). "A smooth nonparametric approach to process capability". *Quality and Reliability Engineering International*. 14:43-48.
- [15] Quandt, R. 1966. "Old and new methods of estimation of the pareto distribution". *Biometrika* 10, 55-82.
- [16] R Development Core Team. (2008). "R: A language and environment for statistical". *R Foundation for Statistical Computing*. Vienna, Austria. ISBN 3-900051-07-0. <http://www.R-project.org/>.
- [17] Somerville, S. & Montgomery, D. (1996). "Process capability indices and non-normal distributions". *Qual. Eng.*, 19(2): 305-316.