Ferreri, Noemí M.[1]
Quaglino, Marta B.[2]
(1) Facultad de Ciencias Exactas, Ingeniería y Agrimensura. UNR
(2) Instituto de Investigaciones Teóricas y Aplicadas. Escuela de Estadística. UNR

PROPIEDADES DISTRIBUCIONALES DE INDICES DE CAPACIDAD DE PROCESOS BAJO DISTRIBUCIONES NO NORMALES

1- INTRODUCCIÓN

El estudio de la capacidad de un proceso, junto con el control estadístico de procesos y el diseño de experimentos son métodos estadísticos que se utilizan desde hace décadas con el propósito de controlar y reducir la variabilidad de procesos y productos, mejorando así calidad y productividad.

Dentro de los estudios de capacidad se utilizan indicadores que cuantifican la habilidad del proceso para cumplir con las especificaciones de clientes tanto internos como externos. Estos son los Índices de Capacidad de Proceso, que se constituyen en los indicadores más utilizados en las empresas para comunicar acerca de la calidad de productos o servicios.

Muchos de los indicadores propuestos en la literatura, entre los que se encuentran los de uso más frecuente en la industria, evalúan varios aspectos de la capacidad de un proceso simultáneamente y el usuario no puede conocer entonces la naturaleza de la falta de capacidad, cuando esta ocurre. Una forma de superar esta limitación es considerar índices que evalúen la capacidad en relación a un único aspecto (variabilidad, centrado, proporción de producción no conforme, etc.) y definir con ellos un vector de capacidad de proceso.

Surge además otro inconveniente relacionado con los índices de capacidad: muchos de ellos han sido construidos suponiendo normalidad de la variable en estudio, de modo que pueden conducir a resultados engañosos si esta no se cumple.

En este trabajo se consideran indicadores que señalan problemas de capacidad en relación a la variabilidad, para procesos con una única variable de calidad continua. Sus objetivos son:

- Presentar diferentes indicadores de capacidad en relación a la variabilidad y analizar sus propiedades poblacionales para procesos con especificaciones bilaterales simétricas[1] y alejamientos de la normalidad.
- Proponer estimadores para los diferentes indicadores de capacidad y estudiar sus propiedades distribucionales por simulación frente a alejamientos de la normalidad.

El trabajo se organiza en 4 secciones de las cuales la Introducción es la primera. En la Sección 2 se presentan diferentes indicadores de capacidad en relación a la variabilidad y se analizan sus propiedades poblacionales; mientras que en la Sección 3 se presentan sus estimadores y se estudian por simulación sus propiedades distribucionales. En la Sección 4 se analizan conjuntamente los resultados obtenidos en las Secciones 2 y 3 y se sustancian las conclusiones más importantes, indicando los aportes que se derivan de este trabajo para el planteo y el estudio de las propiedades de indicadores multidimensionales.

2- INDICADORES DE CAPACIDAD EN RELACION A LA VARIABILIDAD

Cuando la identificación de problemas en la variabilidad del proceso es el objetivo, los índices tienen generalmente la siguiente estructura:

$$\frac{\textit{Variación permitida para el proceso}}{\textit{Variación actual del proceso}}$$

La variación permitida para el proceso (VP) surge a partir de las especificaciones impuestas al producto o servicio y se obtiene haciendo la diferencia entre el límite superior (LSE) y el límite inferior de especificaciones (LIE). La variación actual del proceso (VA) viene dada por la diferencia entre las tolerancias naturales, que son los percentiles superior e inferior de la distribución de la variable en estudio, tales que una proporción dada de dicha distribución se ubica entre ellos.

Un proceso se define como “capaz en relación a la variabilidad” si VA es a lo sumo igual que VP y en ese caso, el cociente toma un valor mayor o igual que 1.

El más divulgado de los índices de este grupo es C_p, presentado por Juran en la década del ’70 del siglo anterior:

[1] Están definidos el límite inferior de especificaciones (LIE) y el límite superior (LSE) y el valor objetivo, T, es el punto medio del intervalo (LIE,LSE)

$$C_p = \frac{LSE - LIE}{6\sigma} \qquad (1),$$

donde σ es la desviación estándar del proceso.

En la expresión de C_p se supone implícitamente que las especificaciones son bilaterales simétricas, 0,00135 es el máximo admitido para la proporción de producción no conforme a cada lado de los límites de especificaciones y la variable en estudio es normal. Si los dos primeros requerimientos se verifican; pero no la normalidad, el valor de C_p puede resultar engañoso.

Chang, Choi y Bai (2002) proponen un indicador para distribuciones asimétricas basado en la idea que en ese caso, la desviación estándar del proceso, σ, puede dividirse en una desviación superior (σ_U) y en una desviación inferior (σ_L) que representan respectivamente el grado de dispersión de los lados superior e inferior respecto de μ. La expresión propuesta es:

$$C^{VP} = C_p \text{mín}\left(\frac{1}{2P_\mu}; \frac{1}{2(1 - P_\mu)}\right) \quad (2),$$

donde C_p *es la expresión (1),* $P_\mu = P(X \leq \mu)$

Otros autores proponen diseñar índices a distribución libre utilizando percentiles. Entre ellos se encuentran Pearn y Chen (1997) que presentan la siguiente expresión:

$$C_{NP} = \frac{LSE - LIE}{P_{99,865} - P_{0,135}} \quad (3),$$

donde $P_{99,865}$ *y* $P_{0,135}$ *son los percentiles correspondientes.*

En síntesis, la expresión (1) solo se puede aplicar bajo normalidad y la expresión (2) se adapta a distribuciones asimétricas, reduciéndose a (1) si la distribución es simétrica. Ambas expresiones están basadas en la desviación estándar. La expresión (3) se basa en percentiles y por lo tanto puede aplicarse para cualquier distribución. Todas las expresiones mencionadas se proponen para especificaciones bilaterales simétricas y suponen un máximo de 0,00135 para la proporción de producción no conforme a cada lado de los límites. Esto último puede modificarse cambiando el valor 6 en (1) o cambiando los percentiles en (3).

Las propiedades poblacionales que se evalúan son la sensibilidad y la especificidad. Una expresión es sensible si advierte sobre la falta de capacidad de los procesos cuando esta se produce; mientras que es específico si no da falsas alarmas cuando el proceso realmente es capaz.

Bajo normalidad, VA coincide con 6σ y la probabilidad de obtener valores menores que μ es 0,50, por lo que cualquiera de las expresiones resultan equivalentes. Además, si el proceso es capaz, VA resulta a lo sumo igual que VP y los índices (1) a (3) toman un valor mayor o igual que 1. Lo contrario ocurre si el proceso no es capaz. Entonces, las tres expresiones resultan sensibles (toman un valor menor que 1 si el proceso no es capaz) y específicas (toman un valor mayor que 1 si el proceso es capaz).

Si el modelo es simétrico pero no normal, las expresiones basadas en σ coinciden. En particular, si se considera una mezcla de distribuciones normales, con el mismo valor esperado pero diferente variabilidad, VA es mayor que 6σ y como consecuencia, estas expresiones sobrevaloran la capacidad del proceso, pudiendo identificar como capaz a un proceso que no lo es. Esto implica que siguen siendo específicas pero no siempre sensibles a la falta de capacidad. Una forma de reducir la pérdida de sensibilidad para estas expresiones es considerar un valor menor que 6 en el denominador, aunque esto significa admitir una mayor proporción de producción no conforme. La expresión (3) no se ve afectada ya que su denominador coincide con VA, resultando siempre sensible y específica.

Si el modelo es asimétrico ninguna de las expresiones propuestas coinciden. Solo (3) resulta sensible y específica; mientras que las restantes son específicas pero pueden resultar no sensibles. En este caso la expresión (2) es la que menos sobrevalora la capacidad. También está la posibilidad de considerar un valor menor que 6, ya mencionada en el párrafo anterior.

La Tabla 1 resume las propiedades poblacionales de las expresiones propuestas. La expresión (3), basada en percentiles resulta siempre sensible y específica, independientemente de la distribución de la variable de calidad. Las restantes pueden resultar no sensibles si el modelo se aleja del normal.

Tabla 1. Sensibilidad y especificidad de los indicadores de capacidad en relación a variabilidad

Distribución para la variable en estudio	*Indicadores de capacidad en relación a la variabilidad*		
	C_p (1)	C^{VP} (2)	C_{NP} (3)
Normal	(▲)	(▲)	(▲)
Mezcla de normales	(○)	(○)	(▲)
Asimétrica	(○)	(○)	(▲)

Referencias: (▲) siempre sensibles y específicas (○) específicas pero no siempre sensibles

La sensibilidad y la especificidad se analizaron suponiendo conocidos la distribución de probabilidad de la variable en estudio y sus parámetros. En la práctica, la capacidad de un proceso se estudia a partir de información muestral y entonces, se requiere el análisis de las propiedades de los estimadores. Esto es lo que se realiza en la siguiente sección.

3- ESTUDIO DE LAS PROPIEDADES DISTRIBUCIONALES DE LOS ESTIMADORES

Para cada uno de los índices de capacidad propuestos en la sección anterior, se define un estimador, por el método de sustitución (plug-in). Los mismos se presentan en la Tabla 2.

Para evaluar las propiedades distribucionales de los estimadores presentados en dicha tabla se consideran los siguientes modelos:

- N(100,10)
- Mezcla de normales, donde el 90 % de los valores corresponden a una distribución N(100,10) y el 10 % restante a una distribución N(100,20)
- Lognormal (2,67; 0,55)[2]

[2] A los valores de esta distribución se les suma la constante 83,20216515 para que E(X) = 100

Tabla 2 Estimadores de los indicadores de capacidad en relación a la variabilidad

Estimadores de las expresiones basadas en σ	*(1)*	$\frac{LSE - LIE}{6S} b$, con b = $\sqrt{\frac{2}{f}} \frac{\Gamma(\frac{f}{2})}{\Gamma(\frac{f-1}{2})}$ y f = n - 1
	(2)	$\min(\frac{LSE - T}{3S(2p_x)}; \frac{T - LIE}{3S(2(1 - p_x))})$
Estimador de la expresión basada en percentiles	*(3)*	$\frac{LSE - LIE}{p_{99,865} - p_{0,135}}$

S: desviación estándar muestral, p_x es la proporción de valores muestrales menores que la media, $p_{0,135}$ y $p_{99,865}$ son los percentiles estimados con el método que R utiliza por omisión y b es un factor de corrección para el sesgo propuesto por Pearn, Kotz y Johnson (1992).

Para las tres distribuciones elegidas, el valor esperado es 100 unidades y la desviación estándar, 10 . Para el tercer modelo, el coeficiente de asimetría es aproximadamente 2.

A su vez, se definen dos casos en relación con las especificaciones, todos ellos correspondientes a especificaciones bilaterales simétricas:

- 1- Especificaciones: (55,145), valor objetivo (T): 100
- 2- Especificaciones: (80,120), valor objetivo (T): 100

De la combinación de los distintos modelos y casos elegidos surgen 6 escenarios diferentes. Si bien tanto los parámetros elegidos para las distribuciones consideradas (media 100 y desviación estándar cercana a 10 para todas las distribuciones) como los límites especificación y valores objetivos son arbitrarios, esto no implica pérdida de generalidad; ya que combinados dan lugar a procesos sin y con problemas de variabilidad, generando un espectro completo de posibles situaciones prácticas (para los modelos considerados en este trabajo, el proceso resulta capaz en relación a la variabilidad en el caso 1 y no capaz en el caso 2).

Para cada uno de los 6 escenarios, se simulan 30000 muestras de tamaños n = 25, 50 y 100. A partir de las distribuciones muestrales empíricas se obtiene el sesgo relativo (SR) y el error cuadrático medio (ECM) de cada uno de los estimadores propuestos. También se analiza la forma de la distribución muestral. Las propiedades mencionadas se definen en la Tabla 3.

Tabla 3 Propiedades de los estimadores de capacidad en relación a la variabilidad

Propiedad	Definición
Sesgo Relativo (SR)	$SR = \left\lvert \frac{E(c) - C}{C} \right\rvert$
Error Cuadrático Medio (ECM)	$ECM = E(c - C)^2 = Var(c) + (E(c) - C)^2$

C es el indicador de capacidad y c, su estimador

El Sesgo Relativo toma valores menores que 0,15 si se considera a los estimadores basados en σ (1 y 2), siendo casi nulo para el estimador 1 si el modelo es el normal. Para el estimador basado en percentiles (3), oscila entre 0,20 y 1, resultando mayor para la mezcla de normales y la distribución asimétrica. Para cualquiera de los estimadores mencionados, los valores de SR se reducen a medida que n aumenta. Estos resultados son similares para los dos casos considerados en este trabajo en relación con las especificaciones (Figura 1). Lo mismo ocurre con el Error Cuadrático Medio, que presenta los valores mayores para el estimador basado en percentiles. Esta diferencia se hace más notoria en el caso 1 (Figura 2).

Figura 1. Sesgo relativo (en valor absoluto) de los estimadores de capacidad en relación a la variabilidad, por modelo, caso y tamaño muestral

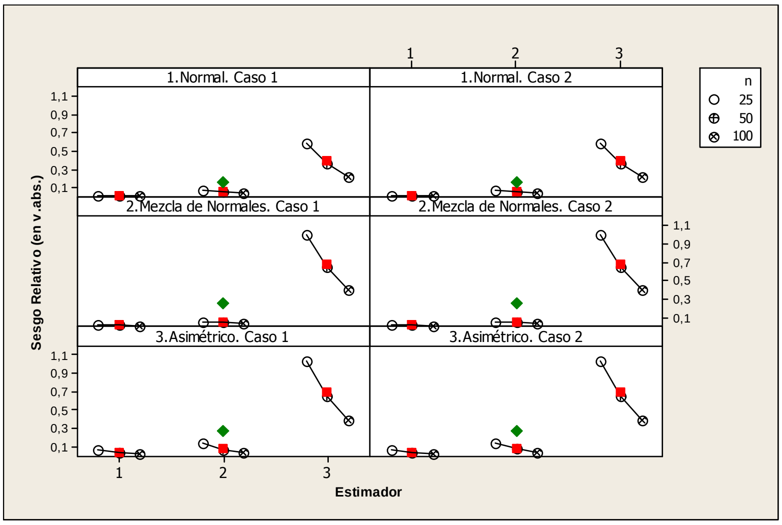
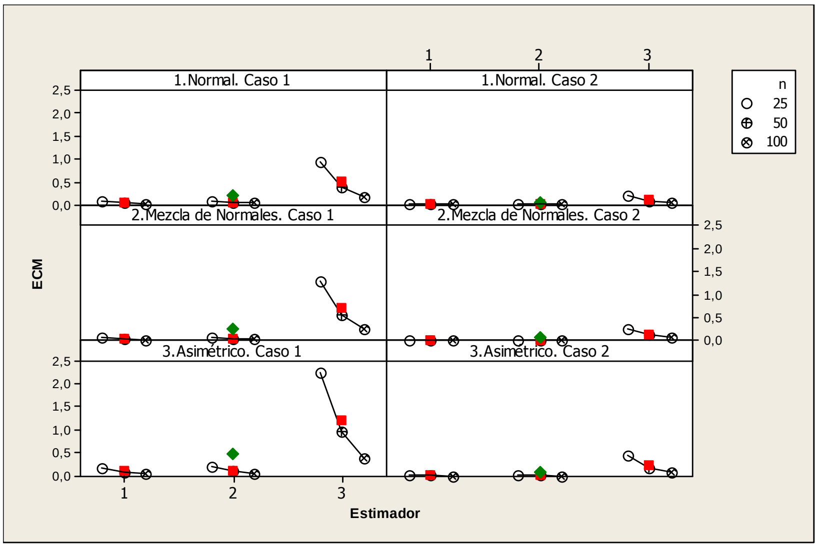


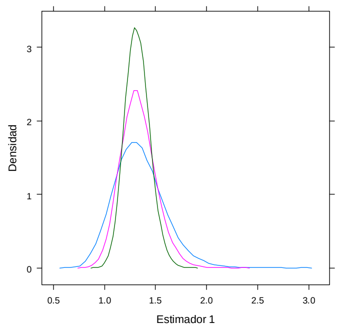
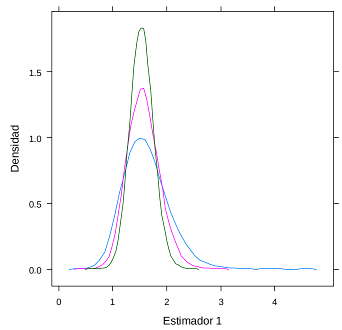
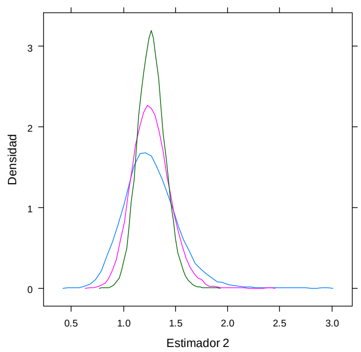
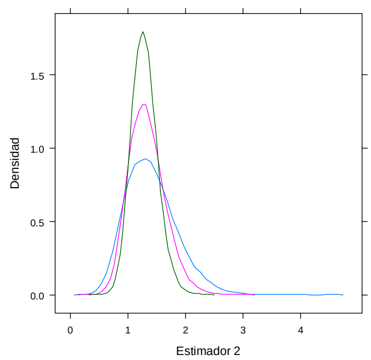
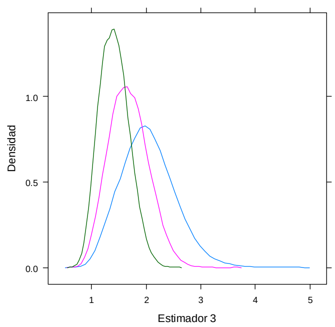
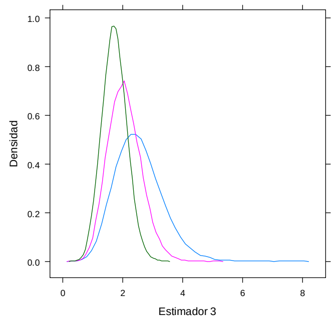
Figura 2. Error Cuadrático Medio de los estimadores de capacidad en relación a la variabilidad, por modelo, caso y tamaño muestral



En la Figura 3 se presentan, a modo de ejemplo, las distribuciones muestrales obtenidas por simulación para la mezcla de normales y para la distribución asimétrica. Como se observa en dichas figuras, las distribuciones de los estimadores se van aproximando a la normal a medida que aumenta el tamaño de muestra. A su vez, las curvas de densidad muestran claramente que el estimador basado en percentiles (3) es sesgado, lo cual es más notorio para el menor tamaño de muestra considerado en este trabajo (n = 25).

En síntesis, los estimadores basados en σ presentan valores de SR y ECM marcadamente inferiores que los correspondientes al estimador basado en percentiles. A su vez, las distribuciones muestrales obtenidas para cualquiera de los estimadores no difiere sustancialmente de la normal para los tamaños muestrales considerados en este trabajo.

Figura 3. Distribuciones muestrales de los estimadores de capacidad en relación a la variabilidad, por modelo, caso y tamaño muestral

	Mezcla de normales	Modelo asimétrico
Estimador 1		
Estimador 2		
Estimador 3		

Referencias: celeste: n = 25; rosa: n = 50; verde: n = 100

4- CONSIDERACIONES FINALES

En este trabajo se propusieron tres indicadores de capacidad en relación a la variabilidad. El primero, C_p , es ampliamente utilizado en la industria y se basa implícitamente en la distribución normal. El segundo, C^{VP} , presenta una modificación para ser utilizada con distribuciones asimétricas y se reduce al primero si la distribución es simétrica. El tercero, C_{NP} , puede ser utilizado con cualquier distribución ya que está basado en percentiles.

Del análisis de la sensibilidad y especificidad de las expresiones en la población surge que solo C_{NP} resulta sensible y específico en todos los casos y modelos considerados. Sin embargo, el estimador propuesto para C_{NP} no presenta buenas propiedades ni siquiera para el mayor tamaño muestral considerado en este trabajo (n = 100), asumiendo valores altos para SR y ECM. En el otro extremo, C_p solo es sensible y específico bajo normalidad y pierde sensibilidad para las mezclas de normales y las distribuciones asimétricas; pero su estimador es el que presenta valores más bajos de SR y ECM.

La alternativa que se propone en este trabajo es considerar como indicador a la expresión C^{VP} y su estimador correspondiente por las siguientes razones: Si el modelo apropiado es el normal, la expresión es sensible y específica; mientras que si se trabaja con distribuciones asimétricas pierde sensibilidad pero en menor medida que C_p. Para mezclas de normales, puede mejorarse su sensibilidad, tomando un valor menor que 6 en el denominador. Además, el estimador propuesto para C^{VP} presenta valores similares o levemente superiores de ECM y SR que el estimador de C_p.

Los indicadores presentados corresponden a la capacidad en relación a la variabilidad, sin embargo podría aplicarse un procedimiento análogo para seleccionar indicadores de capacidad en relación a otros aspectos (centrado, proporción de producción no conforme, etc.), para finalmente definir un indicador multidimensional de capacidad, capaz de identificar simultáneamente distintas dimensiones de capacidad y lo suficientemente flexible como para aplicarse a distintas situaciones. De esta forma, del análisis de la capacidad surgiría no solo la falta de capacidad de un proceso sino también la naturaleza de la misma, permitiendo la planificación de acciones correctivas en forma integral, aun cuando los supuestos clásicos, como la distribución normal de la variable de calidad, no se satisfagan.

REFERENCIAS BIBLIOGRÁFICAS

Chang, Y.; Choi, I.; Bai, D. (2002) "Process Capability Indices for Skewed Populations". *Quality and Reliability Engineering International,* 18, 383-393.

Gunter, B (1989) "The Use and Abuse of Cpk" Parts 1-4, *Quality Progress,* 22, 72-73, 79-80, 86-87, 108-109.

Kotz, S.; Johnson, N. (2002) "Process Capability Indices- A Review, 1992-2000" (con discusión), *Journal of Quality Technology,* 34, 2-19.

Pearn, W.; Chen, K. (1997) "Capability Indices for non-normal distributions with an application in electrolytic capacitor manufacturing", *Microelectronics Reliability, 37,* 1853-1858.

Pearn, W. ; Kotz, S.; Johnson, N. (1992) "Distributional and Inferential Properties of Process Capability Indices", *Journal of Quality Technology,* 24, 216-231.

Pearn, W.; Kotz, S. (2006), Encyclopedia and Handbook of Process Capability Indices. A Comprehensive Exposition of Quality Control Measures. (Series on Quality, Reliability and Engineering Statistics. Vol 12), Singapur: World Scientific.

Shahriari, H.; Abdollahzadeh, M. (2009) "A new multivariate process capability vector", *Quality Engineering,* 21, 290-299.

Wang, F.; Hubele, N.; Lawrence, F.; Miskulin, J.; Shahriari, H. (2000), "Comparison of Three Multivariate Process Capability Indices," *Journal of Quality Technology,* 32, 263-275.