

Intervalos de confianza para los índices de capacidad C_{pm} y C_{pmk} en procesos estacionarios gaussianos

Confidence Intervals for the Capability Indices C_{pm} and C_{pmk} in Stationary Gaussian Processes

RUBÉN DARÍO GUEVARA^{1,a}, JOSÉ ALBERTO VARGAS^{2,b}

¹DEPARTAMENTO DE MATEMÁTICAS Y ESTADÍSTICA, UNIVERSIDAD DEL TOLIMA, IBAGUÉ, COLOMBIA

²DEPARTAMENTO DE ESTADÍSTICA, UNIVERSIDAD NACIONAL DE COLOMBIA, BOGOTÁ

Resumen

Los índices de capacidad suministran una medida numérica del ajuste que tiene un proceso a unos límites de especificación preestablecidos. Los procedimientos existentes para construir intervalos de confianza para los índices de capacidad C_{pm} y C_{pmk} en procesos estacionarios gaussianos muestran bajos porcentajes de cobertura. Este artículo presenta una nueva metodología para construir intervalos de confianza para los índices de capacidad C_{pm} y C_{pmk} , en procesos estacionarios gaussianos y mediante simulación analiza el porcentaje de cobertura para procesos $AR(1)$.

Palabras clave: autocorrelación, análisis de capacidad, índices de capacidad de procesos, estimación.

Abstract

Process capability indices provide a measure of how a process meets pre-set specification limits. Procedures to construct confidence intervals for C_{pm} and C_{pmk} for stationary gaussian processes show low coverage probabilities. This paper presents a new methodology for constructing confidence intervals for the indices C_{pm} and C_{pmk} for stationary gaussian processes and by simulation study the coverage percentage for $AR(1)$ processes is calculated.

Key words: Autocorrelation, Process capability analysis, Process capability indices, Estimation.

^aProfesor. E-mail: rguevara@ut.edu.co

^bProfesor. E-mail: javargasn@unal.edu.co

1. Introducción

Poder identificar qué porcentaje de la producción está por fuera de unas especificaciones es fundamental para que una empresa sea competitiva en un mercado exigente como el actual, y una manera de expresar esta situación es a través de los índices de capacidad de proceso. Un análisis de capacidad clásico supone que el proceso se encuentre bajo control estadístico, que la distribución del proceso sea normal y que exista independencia entre las observaciones. Sin embargo, con relación al último supuesto, muchos procesos presentan cierto grado de autocorrelación distorsionando el significado del índice.

El supuesto de independencia se cuestiona cada día más en ciertos procesos, que bajo la influencia de la tecnología presentan información en tiempo real, y que por su misma naturaleza implican relaciones entre observaciones. Poder realizar análisis de capacidad bajo la situación de autocorrelación permite suministrar a las empresas herramientas confiables para medir su nivel de cumplimiento de estándares externos y facilitar la toma de decisiones con el objetivo de mejorar la calidad de los procesos.

No hay mucha información acerca del análisis de capacidad de procesos en presencia de autocorrelación y en lo concerniente con los índices de capacidad se han encontrado estudios sobre los índices C_p y C_{pk} ; ver por ejemplo las revisiones bibliográficas realizadas por Kotz & Johnson (2002) y Spiring et al. (2003). Zhang (1998) encuentra la varianza de los estimadores de los índices de capacidad C_p y C_{pk} en procesos estacionarios gaussianos y analiza los porcentajes de cobertura de los intervalos de confianza $\widehat{C}_p \pm k\sigma_{\widehat{C}_p}$ y $\widehat{C}_{pk} \pm k\sigma_{\widehat{C}_{pk}}$, con $k = 2$ y $k = 3$, es decir, la proporción en que el verdadero valor de estos índices se halla dentro del intervalo de confianza. Guevara & Vargas (2006) encuentran la varianza de los estimadores de los índices de capacidad C_{pm} y C_{pmk} en procesos estacionarios gaussianos y muestran que los porcentajes de cobertura para los intervalos de confianza $\widehat{C}_{pm} \pm k\sigma_{\widehat{C}_{pm}}$, $\widehat{C}_{pmk} \pm k\sigma_{\widehat{C}_{pmk}}$, con $k = 2, 3$ y $(\widehat{C}_{pm} - k_1\sigma_{\widehat{C}_{pm}}, \widehat{C}_{pm} + k_2\sigma_{\widehat{C}_{pm}})$ y $(\widehat{C}_{pmk} - k_1\sigma_{\widehat{C}_{pmk}}, \widehat{C}_{pmk} + k_2\sigma_{\widehat{C}_{pmk}})$, con $k_1 = 0, 5, \dots, 10$ y $k_2 = 0, 5, \dots, 10$ son bajos.

Este artículo presenta un método basado en regresión lineal, para construir intervalos de confianza de los índices de capacidad C_{pm} y C_{pmk} en procesos estacionarios gaussianos. En particular, para un proceso $AR(1)$, se muestra que los porcentajes de cobertura son altos.

2. Índices de capacidad del proceso

Los índices de capacidad de un proceso suministran información numérica de qué tanto el proceso se ajusta a los límites de especificación establecidos. Existen varios índices de capacidad, pero los más empleados son C_p , C_{pk} , C_{pm} y C_{pmk} , definidos de la siguiente manera:

$$C_p = \frac{USL - LSL}{6\sigma}$$

donde USL y LSL son los límites de especificación superior e inferior respectivamente y σ es la desviación estándar del proceso.

$$\begin{aligned} C_{pk} &= \min\left(\frac{USL - \mu}{3\sigma}, \frac{\mu - LSL}{3\sigma}\right) \\ &= \frac{[a - |\mu - b|]}{3\sigma} \\ &= \frac{[d - |2\mu - m|]}{6\sigma} \end{aligned}$$

donde μ es la media del proceso, $a = (USL - LSL)/2$, $b = (USL + LSL)/2$, $d = (USL - LSL)$, $m = (USL + LSL)$.

$$\begin{aligned} C_{pm} &= \frac{USL - LSL}{6\sqrt{\sigma^2 + (\mu - T)^2}} \\ &= \frac{C_p}{\sqrt{1 + \xi^2}}, \end{aligned}$$

donde T es el valor objetivo y

$$\xi = \frac{\mu - T}{\sigma}$$

$$\begin{aligned} C_{pmk} &= \min\left(\frac{USL - \mu}{3\sqrt{\sigma^2 + (\mu - T)^2}}, \frac{\mu - LSL}{3\sqrt{\sigma^2 + (\mu - T)^2}}\right) \\ &= \frac{C_{pk}}{\sqrt{1 + \xi^2}} \\ &= \frac{a - |\mu - b|}{3\sqrt{\sigma^2 + (\mu - T)^2}} \\ &= \frac{C_{pm}C_{pk}}{C_p} \end{aligned}$$

3. Procesos estacionarios gaussianos

$\{X_t\}$ es un proceso estacionario en el sentido débil si: $\mu_{X(t)}$ es independiente de t y $\gamma_X(t+h, t)$ es independiente de t , para cada h .

Igualmente se define $\{X_t\}$ como un proceso estacionario en el sentido estricto si (X_1, X_2, \dots, X_n) y $(X_{1+h}, X_{2+h}, \dots, X_{n+h})$ tienen la misma distribución conjunta para todo entero h y $n > 0$.

A partir de estas definiciones se puede concluir que todo proceso, estrictamente estacionario con segundo momento finito ($EX_t^2 < \infty$), es estacionario en el sentido débil. Sin embargo, el recíproco no es cierto.

Generalmente en la literatura cuando se habla de un proceso estacionario se hace referencia a un proceso estacionario en el sentido débil (Brockwell & Davis 1996).

El proceso $\{X_t\}$ es una serie de tiempo Gaussiana, si y solo si las funciones de distribución de $\{X_t\}$ son todas normales multivariadas. Por consiguiente, el proceso es estacionario en ambos sentidos (Brockwell & Davis 1990).

Sea $\{X_t\}$ una serie de tiempo estacionaria, la función de autocovarianza (ACVF) de $\{X_t\}$ es (Brockwell & Davis 1996):

$$\gamma_x(h) = Cov(X_{t+h}, X_t)$$

y la función de autocorrelación (ACF) de $\{X_t\}$ es

$$\rho_X(h) = \frac{\gamma_x(h)}{\gamma_x(0)} = Cor(X_{t+h}, X_t)$$

4. Varianza de los estimadores de los índices de capacidad C_{pm} y C_{pmk} en procesos estacionarios gaussianos

Guevara & Vargas (2006) encontraron las siguientes aproximaciones para las varianzas de $\widehat{C_{pm}}$ y $\widehat{C_{pmk}}$ en procesos estacionarios gaussianos:

$$Var(\widehat{C_{pm}}) \approx C_p^2 \left(\frac{\frac{2F(n, \rho_i)}{(n-1)^2} + \frac{4g(n, \rho_i)\xi^2}{n}}{4[f(n, \rho_i) + \xi^2]^3} \right)$$

$$Var(\widehat{C_{pmk}}) \approx C_{pk}^2 \left(\frac{1}{f(n, p_i) + \xi^2} \right) \times \left[\frac{F(n, p_i)}{2(n-1)^2(f(n, p_i) + \xi^2)^2} + \frac{g(n, p_i)}{9n} \left(\frac{1}{C_{pk}} + \frac{6\xi}{2(f(n, p_i) + \xi^2)} \right)^2 \right]$$

donde:

$$\xi = \frac{\mu - T}{\sigma}$$

Se observa que estas varianzas se expresan por funciones de C_p o C_{pk} (para C_{pm} y C_{pmk} respectivamente), el tamaño muestral n , los coeficientes de auto-correlación ρ_i , y ξ .

5. Intervalos de confianza para los índices de capacidad C_{pm} y C_{pmk} en procesos estacionarios gaussianos

En este artículo se propone el siguiente método alternativo para construir intervalos de confianza para C_{pm} y C_{pmk} .

A nivel ilustrativo supóngase que:

1. $n = 50$, $USL = 3$, $LSL = -3$, $\mu = 0$, $T = 1, 2, 3, 4, 5$ y $\sigma = 2$.
2. Se generan 1000 muestras de un proceso $AR(1)$ para cada uno de los siguientes valores de $\phi = 0.15, 0.20, 0.25, \dots, 0.90$.
3. Para cada ϕ se calculan los índices de capacidad C_{pm} y C_{pmk} , y las expresiones $\max(\widehat{C_{pm_i}} - C_{pm})$ y $\max|\widehat{C_{pmk_i}} - C_{pmk}|$ para $i = 1, 2, \dots, 1000$.

Mediante un diagrama de dispersión se observa que existe un comportamiento exponencial entre ϕ y $\max(\widehat{C_{pm_i}} - C_{pm})$ (ver figura 1). Comportamiento semejante existe entre ϕ y $\max|\widehat{C_{pmk_i}} - C_{pmk}|$ (ver figura 3).

Empleando la transformación $\ln(y) = \ln(\beta_0) + \beta_1\phi$, donde:

- ϕ representa el coeficiente autorregresivo, y
- $y = \max(\widehat{C_{pm_i}} - C_{pm})$ o $y = \max|\widehat{C_{pmk_i}} - C_{pmk}|$, según sea el caso, se genera una relación lineal entre ϕ y $\ln(y)$ (ver figuras 2 y 4).

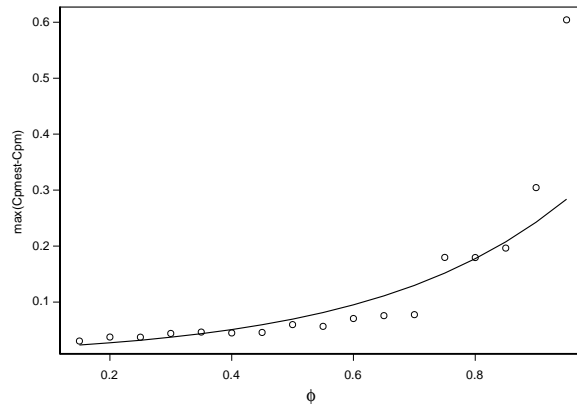


FIGURA 1: Relación exponencial entre ϕ y $\max(\widehat{C_{pm_i}} - C_{pm})$.

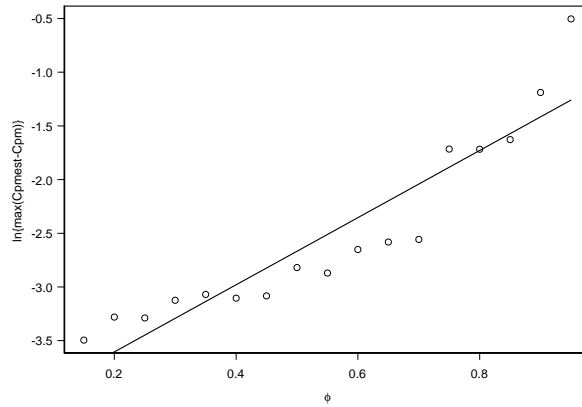


FIGURA 2: Relación lineal entre ϕ y $\ln [\max (\widehat{C}_{pm} - C_{pm})]$.

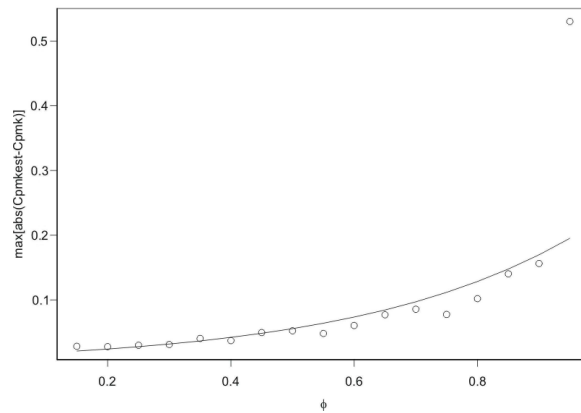


FIGURA 3: Relación exponencial entre ϕ y $\max |\widehat{C}_{pmk} - C_{pmk}|$.

Mediante regresión lineal se obtienen los estimadores de los coeficientes β_0 y β_1 . En la tabla 1 se presentan algunos valores obtenidos para β_0 y β_1 , que dependen de USL, LSL, μ, σ, T y n .

Se construye un intervalo de confianza de la forma $\widehat{C}_{pm} \pm \widehat{y}$ o $\widehat{C}_{pmk} \pm \widehat{y}$, donde $\widehat{y} = \exp(\widehat{\beta}_0 + \widehat{\beta}_1\phi)$ y representa $k\widehat{\sigma}_{c_{pm}}$ o $k\widehat{\sigma}_{c_{pmk}}$, según sea el caso. Dado que se conoce $\widehat{\sigma}_{c_{pm}}$ (o $\widehat{\sigma}_{c_{pmk}}$), al dividir \widehat{y} por $\widehat{\sigma}_{c_{pm}}$ (o $\widehat{\sigma}_{c_{pmk}}$) se obtiene un valor estimado para k , el cual varía de una muestra a otra. El k obtenido para cada muestra es variable porque depende de ϕ y de $\widehat{\sigma}_{c_{pm}}$ (o $\widehat{\sigma}_{c_{pmk}}$).

Mediante simulación se encuentran los porcentajes de cobertura; en particular presentamos los resultados de la simulación para procesos $AR(1)$ con $\mu = 0, \sigma = 2$ y $\phi = 0.75$. Para ello se generaron 1000 muestras de tamaño $n = 50$, para los valores objetivos $T = 1, 2, 3, 4$ y 5 , y se construyeron intervalos de confianza de la

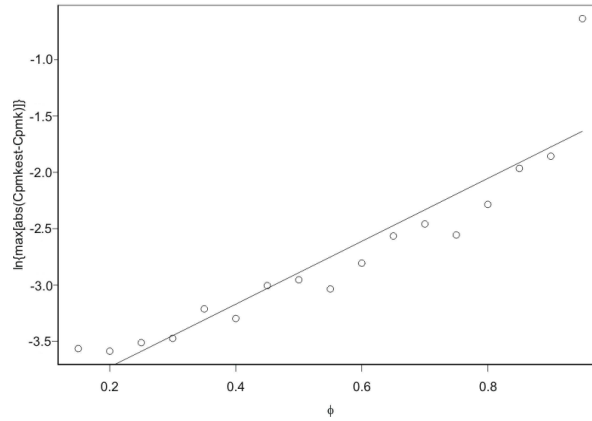


FIGURA 4: Relación lineal entre ϕ y $\ln(\max |\widehat{C}_{pmk} - C_{pmk}|)$.

forma $\widehat{C}_{pm} \pm \widehat{y}$ o $\widehat{C}_{pmk} \pm \widehat{y}$, con $\widehat{y} = \exp(\widehat{\beta}_0 + \widehat{\beta}_1 \phi)$, donde $\widehat{\beta}_0$ y $\widehat{\beta}_1$ son los valores estimados de β_0 y β_1 , que toman los valores que se presentan en la tabla 1.

TABLA 1: Valores estimados de β_0 y β_1 para los intervalos de confianza de C_{pm} y C_{pmk} construidos mediante regresión, para procesos $AR(1)$, con $USL = 3$, $LSL = -3$, $n = 50$, $\mu = 0$, $\sigma = 2$.

T	C_{pm}		C_{pmk}	
	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_0$	$\widehat{\beta}_1$
5	-4.35	2.83	-4.29	2.79
4	3.83	2.70	-4.05	2.55
3	-3.11	2.58	-3.68	2.26
2	-2.00	1.55	-2.35	1.03
1	-0.74	0.35	-1.03	0.44

En la tabla 2 se observa que el porcentaje de cobertura del índice C_{pm} bajo el modelo de regresión es mayor o igual al 98.6%, es decir, de los 1000 intervalos construidos mediante regresión, 986 contienen el verdadero valor del índice; mientras que el porcentaje de cobertura para los intervalos de la forma $\widehat{C}_{pm} \pm k\sigma_{\widehat{C}_{pm}}$ oscila entre 68.3% y 90.6% cuando $k = 2$ y entre 84.2% y 97.5% cuando $k = 3$ (Guevara 2005). Para C_{pmk} el porcentaje de cobertura es mayor o igual al 88.2% cuando se emplean intervalos de confianza construidos con el modelo de regresión; estos porcentajes de cobertura oscilan entre 55.3% y 67.4% y entre 69.9% y 78.2% cuando se emplean intervalos de la forma $\widehat{C}_{pmk} \pm k\sigma_{\widehat{C}_{pmk}}$, con $k = 2, 3$ (Guevara 2005).

Se debe resaltar que los datos que se presentan en las tablas 1 y 2 corresponden a la situación planteada al principio de esta sección; sin embargo, la metodología se puede llevar a cualquier otra situación. En el apéndice se expone un algoritmo para encontrar las estimaciones de los coeficientes de regresión para C_{pm} y C_{pmk} cuando se conocen los límites de especificación, la media y la desviación estándar

del proceso, el tamaño de la muestra y el valor objetivo. Un segundo algoritmo muestra cómo construir los intervalos de confianza cuando se tiene una muestra en particular.

TABLA 2: Porcentajes de cobertura para el intervalo de confianza de los índices C_{pm} y C_{pmk} contruidos mediante regresión *vs.* intervalos de confianza de la forma $\widehat{C}_{pm} \pm k\sigma_{\widehat{C}_{pm}}$, $\widehat{C}_{pmk} \pm k\sigma_{\widehat{C}_{pmk}}$, con $k = 2, 3$, para procesos $AR(1)$, con $n = 50$, $USL = 3$, $LSL = -3$, $\mu = 0$, $\sigma = 2$ y $\phi = 0.75$.

T	C_{pm}			C_{pmk}		
	Regresión	\widehat{C}_{pm}	\widehat{C}_{pm}	Regresión	\widehat{C}_{pmk}	\widehat{C}_{pmk}
		$\pm 2\sigma_{\widehat{C}_{pm}}$	$\pm 3\sigma_{\widehat{C}_{pm}}$		$\pm 2\sigma_{\widehat{C}_{pmk}}$	$\pm 3\sigma_{\widehat{C}_{pmk}}$
5	98.6 %	68.3 %	84.2 %	92.2 %	55.3 %	69.9 %
4	99.9 %	72.5 %	86.9 %	88.2 %	59.1 %	71.3 %
3	100.0 %	78.6 %	91.0 %	88.0 %	62.0 %	73.5 %
2	100.0 %	83.7 %	93.9 %	95.1 %	62.5 %	74.5 %
1	100.0 %	90.6 %	97.5 %	100.0 %	67.4 %	78.2 %

6. Conclusiones

Para procesos estacionarios gaussianos con modelos $AR(1)$ hemos mostrado un método que permite construir intervalos de confianza para los índices de capacidad C_{pm} y C_{pmk} . En el caso de C_{pm} este método aprovecha la relación exponencial que existe entre ϕ y la máxima diferencia que se presenta entre el índice de capacidad y sus estimaciones. Para C_{pmk} se tiene en cuenta la relación exponencial que existe entre ϕ y la máxima diferencia absoluta entre el índice de capacidad y sus estimaciones. A través de una transformación de estas relaciones se construye una relación lineal entre las observaciones, se aplican procesos de regresión lineal y mediante simulación se encuentran los valores estimados de los coeficientes de regresión que permiten construir el intervalo de confianza para cada índice.

Mediante simulación se encontró que los porcentajes de cobertura que presentan estos intervalos son significativamente altos, situación que no ocurren en los métodos tradicionales existentes.

Recibido: abril de 2006

Aceptado: septiembre de 2006

Referencias

- Brockwell, P. & Davis, R. (1990), *Time Series: Theory and Methods*, Springer, New York.
- Brockwell, P. & Davis, R. (1996), *Introduction to Time Series and Forecasting*, Springer, New York.

- Guevara, R. D. (2005), Comparación de los índices de capacidad C_p , C_{pk} , C_{pm} C_{pmk} para datos autocorrelacionados, Tesis de maestría, Universidad Nacional de Colombia, Bogotá.
- Guevara, R. D. & Vargas, J. A. (2006), Comparison of Process Capability Indices Under Autocorrelated Data. Enviado para publicación en Journal of Applied Statistics.
- Kotz, S. & Johnson, N. (2002), 'Process Capability Indices -A Review, 1992-2000', *Journal of Quality Technology* **34**, 2-19.
- Spiring, F., Leung, B., Cheng, S. & Yeung, A. (2003), 'A Bibliography of Process Capability Papers', *Quality and Reliability Engineering International* **19**, 445-460.
- Zhang, N. F. (1998), 'Estimating Process Capability Indexes for Autocorrelated Data', *Journal of Applied Statistics* **25**, 559-574.

Apéndice

Algoritmo para calcular los coeficientes de regresión $\beta_{0C_{pm}}$, $\beta_{1C_{pm}}$, $\beta_{0C_{pmk}}$, $\beta_{1C_{pmk}}$, en procesos estacionarios gaussianos, **AR(1)**

Función CalcB()

Leer el tamaño de la muestra, n

Leer límites de especificación, LSL, USL

Leer valor objetivo, T

Leer media del proceso, μ

Leer desviación estándar del proceso, σ

Para ϕ de 0.15, 0.2, 0.25, ..., 0.90

Calcular $C_{pm\phi}$ y $C_{pmk\phi}$

$DifMaxC_{pm\phi} \leftarrow 0$

$DifMaxC_{pmk\phi} \leftarrow 0$

Para muestra i : de 1 a 1000

Calcular \widehat{C}_{pm_i}

Calcular \widehat{C}_{pmk_i}

$DifMaxC_{pm\phi} \leftarrow \max(\widehat{C}_{pm_i} - C_{pm\phi}, DifMaxC_{pm\phi})$

$DifMaxC_{pmk\phi} \leftarrow \max(\widehat{C}_{pmk_i} - C_{pmk\phi}, DifMaxC_{pmk\phi})$

Fin Para

$$y_\phi \leftarrow \ln(\text{DifMaxCpm}_\phi)$$

$$w_\phi \leftarrow \ln(\text{DifMaxCpmk}_\phi)$$

Fin Para

Obtener por regresión lineal los estimadores de los coeficientes de regresión $\beta_{o_{C_{pm}}}$, $\beta_{1_{C_{pm}}}$, $\beta_{o_{C_{pmk}}}$, $\beta_{1_{C_{pmk}}}$ empleando los modelos $\ln(y_\phi) = \ln(\beta_{o_{C_{pm}}}) + \beta_{1_{C_{pm}}} \phi$ y $\ln(w_\phi) = \ln(\beta_{o_{C_{pmk}}}) + \beta_{1_{C_{pmk}}} \phi$.

Fin Función

Algoritmo para construir los intervalos de confianza para los índices de capacidad C_{pm} y C_{pmk} , en procesos estacionarios gaussianos $AR(1)$, empleando regresión

Función *Calc-IC-C_{pm}-C_{pmk}* ($\beta_{o_{C_{pm}}}$, $\beta_{1_{C_{pm}}}$, $\beta_{o_{C_{pmk}}}$, $\beta_{1_{C_{pmk}}}$, LSL , USL , μ , σ , T , n , ϕ)

Leer los datos de la muestra

Calcular \widehat{C}_{pm}

Calcular \widehat{C}_{pmk}

$$IC_{C_{pm}} = \widehat{C}_{pm} \pm \exp\left(\widehat{\beta_{o_{C_{pm}}}} + \widehat{\beta_{1_{C_{pm}}}} \phi\right)$$

$$IC_{C_{pmk}} = \widehat{C}_{pmk} \pm \exp\left(\widehat{\beta_{o_{C_{pmk}}}} + \widehat{\beta_{1_{C_{pmk}}}} \phi\right)$$

Fin Función