

UNIVERSIDAD DE GRANADA



**UNIVERSIDAD  
DE GRANADA**

NUEVAS APORTACIONES Y ANÁLISIS  
DE LA ESTIMACIÓN DEL ÍNDICE DE  
CAPACIDAD DE UN PROCESO.

TESIS DOCTORAL

Directores:

Prof. Dr. D. Juan Francisco Muñoz Rosas  
Prof. Dr. D<sup>a</sup>. Encarnación Álvarez Verdejo

Doctorando:

Pablo José Moya Fernández

DEPARTAMENTO DE MÉTODOS CUANTITATIVOS PARA LA  
ECONOMÍA Y LA EMPRESA

Programa de Doctorado en Ciencias Económicas y Empresariales

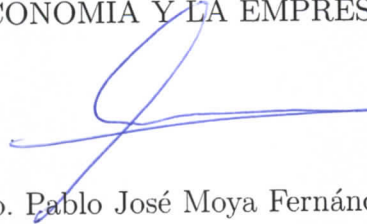
Granada 2017

Editor: Universidad de Granada. Tesis Doctorales  
Autor: Pablo José Moya Fernández  
ISBN: 978-84-9163-648-9  
URI: <http://hdl.handle.net/10481/48757>

# NUEVAS APORTACIONES Y ANÁLISIS DE LA ESTIMACIÓN DEL ÍNDICE DE CAPACIDAD DE UN PROCESO.

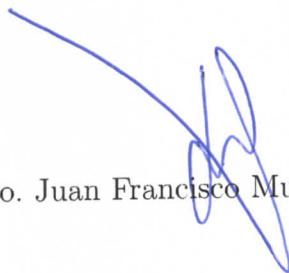
Memoria presentada por Pablo José Moya Fernández para aspirar al Título de  
Doctor con Mención Internacional por la Universidad de Granada

DEPARTAMENTO DE MÉTODOS CUANTITATIVOS PARA LA  
ECONOMÍA Y LA EMPRESA




Fdo. Pablo José Moya Fernández

Vº Bº Directores de tesis:



Fdo. Juan Francisco Muñoz Rosas



Fdo. Encarnación Álvarez Verdejo

UNIVERSIDAD DE GRANADA

Granada 2017

El doctorando / The *doctoral candidate* [ **Pablo José Moya Fernández** ] y los directores de la tesis / and the thesis supervisor/s: [ **Juan Francisco Muñoz Rosas y Encarnación Álvarez Verdejo** ]

Garantizamos, al firmar esta tesis doctoral, que el trabajo ha sido realizado por el doctorando bajo la dirección de los directores de la tesis y hasta donde nuestro conocimiento alcanza, en la realización del trabajo, se han respetado los derechos de otros autores a ser citados, cuando se han utilizado sus resultados o publicaciones.

/

*Guarantee, by signing this doctoral thesis, that the work has been done by the doctoral candidate under the direction of the thesis supervisor/s and, as far as our knowledge reaches, in the performance of the work, the rights of other authors to be cited (when their results or publications have been used) have been respected.*

Lugar y fecha / Place and date:

Granada, 14 de junio de 2017

Director/es de la Tesis / *Thesis supervisor/s*;

Doctorando / *Doctoral candidate*:

Firma / Signed

Firma / Signed



## AGRADECIMIENTOS

No me gustaría dejar pasar la oportunidad de dar las gracias a mis directores de tesis, la Dra. Encarnación Álvarez y el Dr. Juan Francisco Muñoz, cuyas aportaciones y orientación han hecho posible la presente tesis doctoral. Gracias por la confianza que habéis depositado en mí y por haber puesto a mi alcance los medios necesarios para llevar a cabo las investigaciones durante todo el proceso. No tengo palabras para expresar vuestra generosidad.

Les estoy y estaré enormemente agradecido por haberme enseñado a trabajar e investigar y espero seguir mejorando estas aptitudes junto a ellos. Es para mí un honor haber podido contar con su inestimable ayuda, sin la que este trabajo no hubiese visto la luz.

Gracias a mis compañeros del Departamento de Métodos Cuantitativos para la Economía y la Empresa por su apoyo.

En último lugar, que no en último puesto, este trabajo está dedicado a mi familia, mis padres y hermanos, por su ejemplo; y a mi pareja, por su apoyo y paciencia a lo largo del camino y por animarme a pensar a lo grande.



# Índice general

<b>1. Introducción</b>	<b>1</b>
1.1. El control estadístico de calidad . . . . .	1
1.1.1. Gráficos de control . . . . .	7
1.2. Análisis de la capacidad de un proceso . . . . .	11
1.3. Objetivos . . . . .	14
<b>2. Análisis del índice de capacidad en distintos escenarios</b>	<b>21</b>
2.1. Introducción . . . . .	21
2.2. Caso de una única muestra . . . . .	22
2.2.1. Supuestos básicos . . . . .	24
2.2.2. Estimación puntual y por intervalos de confianza . . . . .	40
2.2.3. Estudio de simulación Monte Carlo . . . . .	49
2.2.4. Conclusiones . . . . .	73
2.3. Extensión al caso de m-muestras . . . . .	74
2.3.1. Estimadores de la desviación típica y del índice de capacidad . . . . .	75
2.3.2. Estudio de simulación Monte Carlo . . . . .	77
2.3.3. Conclusiones . . . . .	90
<b>3. Aportaciones relacionadas con la función <math>c_4[\cdot]</math>.</b>	<b>103</b>
3.1. Introducción . . . . .	103
3.2. Definición y justificación de la función $c_4[\cdot]$ . . . . .	106
3.2.1. Análisis de la distribución de la cuasivarianza muestral en poblaciones finitas e infinitas . . . . .	111
3.3. Nuevas expresiones para el cálculo de la función $c_4[\cdot]$ . . . . .	112
3.4. Conclusiones . . . . .	124
<b>4. A new R-package: ‘Capability’</b>	<b>127</b>
4.1. Introduction. The statistical R software . . . . .	127
4.2. The package: ‘Capability’ . . . . .	130
<b>5. Conclusiones, limitaciones y futuras líneas de investigación</b>	<b>201</b>



5.1. Conclusiones . . . . .	201
5.2. Limitaciones y Futuras líneas de investigación . . . . .	208
<b>Bibliografía</b>	<b>213</b>

# Índice de figuras

1.1.	Gráfico de Control para la media de un proceso que se encuentra bajo control. Datos reales tomados de la base de datos “piston-rings” incluidos en la librería “qcc” (Scrucca 2004) del software estadístico R. . . . .	9
1.2.	Relación entre límites de especificación y límites de tolerancia natural según el <i>PCI</i> . . . . .	13
1.3.	Diferencias entre $C_p$ y $C_{pk}$ en función del valor de $k$ y para unos determinados límites de especificación. . . . .	14
2.1.	Gráfico de control para la media. Se seleccionan muestras de una distribución Normal $\mu = 10$ y $\sigma = 1$ . La producción con incidencia representa una proporción $p^* = 0,2$ , y su característica de calidad está afectada por los valores $\epsilon_\mu = 1$ y $\epsilon_\sigma = 0$ . . . . .	29
2.2.	Gráfico de control para la desviación típica. Se seleccionan muestras de una distribución Normal $\mu = 10$ y $\sigma = 1$ . La producción con incidencia representa una proporción $p^* = 0,2$ , y su característica de calidad está afectada por los valores $\epsilon_\mu = 1$ y $\epsilon_\sigma = 0$ . . . . .	30
2.3.	Gráfico de control para la media. Se seleccionan muestras de una distribución Normal $\mu = 10$ y $\sigma = 1$ . La producción con incidencia representa una proporción $p^* = 0,2$ , y su característica de calidad está afectada por los valores $\epsilon_\mu = 2$ y $\epsilon_\sigma = 0$ . . . . .	31
2.4.	Gráfico de control para la desviación típica. Se seleccionan muestras de una distribución Normal $\mu = 10$ y $\sigma = 1$ . La producción con incidencia representa una proporción $p^* = 0,2$ , y su característica de calidad está afectada por los valores $\epsilon_\mu = 2$ y $\epsilon_\sigma = 0$ . . . . .	32
2.5.	Gráfico de control para la media. Se seleccionan muestras de una distribución Normal $\mu = 10$ y $\sigma = 1$ . La producción con incidencia representa una proporción $p^* = 0,2$ , y su característica de calidad está afectada por los valores $\epsilon_\mu = 0$ y $\epsilon_\sigma = 1$ . . . . .	34

2.6.	Gráfico de control para la desviación típica. Se seleccionan muestras de una distribución Normal $\mu = 10$ y $\sigma = 1$ . La producción con incidencia representa una proporción $p^* = 0,2$ , y su característica de calidad está afectada por los valores $\epsilon_\mu = 0$ y $\epsilon_\sigma = 1$ .	35
2.7.	Gráfico de control para la media. Se seleccionan muestras de una distribución Normal $\mu = 10$ y $\sigma = 1$ . La producción con incidencia representa una proporción $p^* = 0,2$ , y su característica de calidad está afectada por los valores $\epsilon_\mu = 0$ y $\epsilon_\sigma = 2$ .	36
2.8.	Gráfico de control para la desviación típica. Se seleccionan muestras de una distribución Normal $\mu = 10$ y $\sigma = 1$ . La producción con incidencia representa una proporción $p^* = 0,2$ , y su característica de calidad está afectada por los valores $\epsilon_\mu = 0$ y $\epsilon_\sigma = 2$ .	37
2.9.	Gráfico de control para la media. Se seleccionan muestras de una distribución Normal $\mu = 10$ y $\sigma = 1$ . La producción con incidencia representa una proporción $p^* = 0,2$ , y su característica de calidad está afectada por los valores $\epsilon_\mu = 1$ y $\epsilon_\sigma = 1$ .	38
2.10.	Gráfico de control para la desviación típica. Se seleccionan muestras de una distribución Normal $\mu = 10$ y $\sigma = 1$ . La producción con incidencia representa una proporción $p^* = 0,2$ , y su característica de calidad está afectada por los valores $\epsilon_\mu = 1$ y $\epsilon_\sigma = 1$ .	39
2.11.	Valores medios del cuadrado del error estándar relativo de los estimadores $\widehat{C}_{p,c4}$ y $\widehat{C}_p$ , y basados en datos procedentes de las distribuciones Normal, Gamma y Uniforme. También se incluyen escenarios con proporciones del 5% y del 20% de valores atípicos en la muestra.	44
2.12.	Valores de las diferencias $D1$ y $D2$ para cuantiles de órdenes usados en la construcción de los intervalos de confianza con niveles de significación $\alpha = \{0,01, 0,05, 0,1\}$ y tamaños muestrales comprendidos entre 10 y 200.	47
2.13.	Valores de $RB$ (%) y $RRMSE$ (%) de los estimadores de $\sigma = 1$ ( $SD$ ) y $C_p = C_{pk} = 1$ ( $PCI$ ), y $Coverage$ (%) de los intervalos de confianza para $C_p = C_{pk} = 1$ ( $PCI$ ). Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan $N = 500$ datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.	56

- 2.14. Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,25$ , y  $Coverage$  (%) de los intervalos de confianza para  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,25$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 500$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . . 57
- 2.15. Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,5$ , y  $Coverage$  (%) de los intervalos de confianza para  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,5$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 500$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . . 58
- 2.16. Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_p = C_{pk} = 1$  ( $PCI$ ), y  $Coverage$  (%) de los intervalos de confianza para  $C_p = C_{pk} = 1$  ( $PCI$ ). Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 500$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . . 59
- 2.17. Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,25$ , y  $Coverage$  (%) de los intervalos de confianza para  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,25$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 500$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . . 60
- 2.18. Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,5$ , y  $Coverage$  (%) de los intervalos de confianza para  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,5$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 500$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . . 61

2.19. Valores de $RB$ (%) y $RRMSE$ (%) de los estimadores de $\sigma = 1$ ( $SD$ ) y $C_p = C_{pk} = 1$ ( $PCI$ ), y $Coverage$ (%) de los intervalos de confianza para $C_p = C_{pk} = 1$ ( $PCI$ ). Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan $N = 10000$ datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . .	62
2.20. Valores de $RB$ (%) y $RRMSE$ (%) de los estimadores de $\sigma = 1$ ( $SD$ ) y $C_{pk} = 1$ ( $PCI$ ), con $k = 0,25$ , y $Coverage$ (%) de los intervalos de confianza para $C_{pk} = 1$ ( $PCI$ ), con $k = 0,25$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan $N = 10000$ datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . .	63
2.21. Valores de $RB$ (%) y $RRMSE$ (%) de los estimadores de $\sigma = 1$ ( $SD$ ) y $C_{pk} = 1$ ( $PCI$ ), con $k = 0,5$ , y $Coverage$ (%) de los intervalos de confianza para $C_{pk} = 1$ ( $PCI$ ), con $k = 0,5$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan $N = 10000$ datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . .	64
2.22. Valores de $RB$ (%) y $RRMSE$ (%) de los estimadores de $\sigma = 1$ ( $SD$ ) y $C_p = C_{pk} = 1$ ( $PCI$ ), y $Coverage$ (%) de los intervalos de confianza para $C_p = C_{pk} = 1$ ( $PCI$ ). Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan $N = 10000$ datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . .	65
2.23. Valores de $RB$ (%) y $RRMSE$ (%) de los estimadores de $\sigma = 1$ ( $SD$ ) y $C_{pk} = 1$ ( $PCI$ ), con $k = 0,25$ , y $Coverage$ (%) de los intervalos de confianza para $C_{pk} = 1$ ( $PCI$ ), con $k = 0,25$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan $N = 10000$ datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . .	66

2.24. Valores de $RB$ (%) y $RRMSE$ (%) de los estimadores de $\sigma = 1$ ( $SD$ ) y $C_{pk} = 1$ ( $PCI$ ), con $k = 0,5$ , y $Coverage$ (%) de los intervalos de confianza para $C_{pk} = 1$ ( $PCI$ ), con $k = 0,5$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan $N = 10000$ datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . .	67
2.25. Valores de $RB$ (%) y $RRMSE$ (%) de los estimadores de $\sigma = 1$ ( $SD$ ) y $C_p = C_{pk} = 1$ ( $PCI$ ), y $Coverage$ (%) de los intervalos de confianza para $C_p = C_{pk} = 1$ ( $PCI$ ). Se seleccionan los datos muestrales de una distribución Gamma (población infinita). Para la población finita, se seleccionan $N = 500$ datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . .	68
2.26. Valores de $RB$ (%) y $RRMSE$ (%) de los estimadores de $\sigma = 1$ ( $SD$ ) y $C_p = C_{pk} = 1$ ( $PCI$ ), y $Coverage$ (%) de los intervalos de confianza para $C_p = C_{pk} = 1$ ( $PCI$ ). Se seleccionan los datos muestrales de una distribución Uniforme (población infinita). Para la población finita, se seleccionan $N = 500$ datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . .	69
2.27. Valores de $RB$ (%) y $RRMSE$ (%) de los estimadores de $\sigma = 0,01291$ ( $SD$ ) y $C_p = C_{pk} = 1$ ( $PCI$ ), y $Coverage$ (%) de los intervalos de confianza para $C_p = C_{pk} = 1$ ( $PCI$ ). Se considera el conjunto de datos pistonrings como población, con $N = 125$ y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple con y sin reemplazamiento. . . . .	70
2.28. Valores de $RB$ (%) de los estimadores $\hat{C}_{p,R}$ (usando los rangos muestrales y la constante $d_2$ ), $\hat{C}_{p,R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\hat{C}_{p,S}$ (usando la desviación típica muestral) y $\hat{C}_{p,P}$ (usando la desviación típica ponderada) de $C_p = 1$ . Datos seleccionados de la distribución Normal con $\sigma = 1$ . . . . .	80
2.29. Valores de $RB$ (%) de los estimadores $\hat{\sigma}_R$ (usando los rangos muestrales y la constante $d_2$ ), $\hat{\sigma}_{R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\hat{\sigma}_S$ (usando la desviación típica muestral) y $\hat{\sigma}_P$ (usando la desviación típica ponderada) de $\sigma = 1$ . Datos seleccionados de la distribución Normal. . . . .	81

2.30. Valores de $RB$ (%) de los estimadores $\widehat{C}_{p.R}$ (usando los rangos muestrales y la constante $d_2$ ), $\widehat{C}_{p.R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\widehat{C}_{p.S}$ (usando la desviación típica muestral) y $\widehat{C}_{p.P}$ (usando la desviación típica ponderada) de $C_p = 1$ . Datos seleccionados de la distribución Gamma con $\sigma = 1$ .	82
2.31. Valores de $RB$ (%) de los estimadores $\widehat{\sigma}_R$ (usando los rangos muestrales y la constante $d_2$ ), $\widehat{\sigma}_{R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\widehat{\sigma}_S$ (usando la desviación típica muestral) y $\widehat{\sigma}_P$ (usando la desviación típica ponderada) de $\sigma = 1$ . Datos seleccionados de la distribución Gamma.	83
2.32. Valores de $RB$ (%) de los estimadores $\widehat{C}_{p.R}$ (usando los rangos muestrales y la constante $d_2$ ), $\widehat{C}_{p.R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\widehat{C}_{p.S}$ (usando la desviación típica muestral) y $\widehat{C}_{p.P}$ (usando la desviación típica ponderada) de $C_p = 1$ . Datos seleccionados de la distribución Uniforme con $\sigma = 1$ .	84
2.33. Valores de $RB$ (%) de los estimadores $\widehat{\sigma}_R$ (usando los rangos muestrales y la constante $d_2$ ), $\widehat{\sigma}_{R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\widehat{\sigma}_S$ (usando la desviación típica muestral) y $\widehat{\sigma}_P$ (usando la desviación típica ponderada) de $\sigma = 1$ . Datos seleccionados de la distribución Uniforme.	85
2.34. Valores de $RB$ (%) de los estimadores $\widehat{C}_{pk.R}$ (usando los rangos muestrales y la constante $d_2$ ), $\widehat{C}_{pk.R2}$ (usando los rangos muestrales y las constantes $d_2$ and $d_2^*$ ), $\widehat{C}_{pk.S}$ (usando la desviación típica muestral) y $\widehat{C}_{pk.P}$ (usando la desviación típica ponderada) de $C_{pk} = 1$ . Datos seleccionados de la distribución Normal con $\sigma = 1$ en un proceso no centrado.	86
2.35. Valores de $RB$ (%) de los estimadores $\widehat{\sigma}_R$ (usando los rangos muestrales y la constante $d_2$ ), $\widehat{\sigma}_{R2}$ (usando los rangos muestrales y las constantes $d_2$ and $d_2^*$ ), $\widehat{\sigma}_S$ (usando la desviación típica muestral) y $\widehat{\sigma}_P$ (usando la desviación típica ponderada) de $\sigma = 1$ . Datos seleccionados de la distribución Normal en un proceso no centrado.	87
2.36. Valores de $RB$ (%) de los estimadores $\widehat{C}_{p.R}$ (usando los rangos muestrales y la constante $d_2$ ), $\widehat{C}_{p.R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\widehat{C}_{p.S}$ (usando la desviación típica muestral) y $\widehat{C}_{p.P}$ (usando la desviación típica ponderada) de $C_p = 1$ . Datos seleccionados de la distribución Gamma con $\sigma = 3$ .	88

2.37. Valores de $RB$ (%) de los estimadores $\hat{\sigma}_R$ (usando los rangos muestrales y la constante $d_2$ ), $\hat{\sigma}_{R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\hat{\sigma}_S$ (usando la desviación típica muestral) y $\hat{\sigma}_P$ (usando la desviación típica ponderada) de $\sigma = 1$ . Datos seleccionados de la distribución Gamma. . . . .	89
2.38. Valores de $RRMSE$ (%) de los estimadores $\hat{C}_{p,R}$ (usando los rangos muestrales y la constante $d_2$ ), $\hat{C}_{p,R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\hat{C}_{p,S}$ (usando la desviación típica muestral) y $\hat{C}_{p,P}$ (usando la desviación típica ponderada) de $C_p = 1$ . Datos seleccionados de la distribución Normal con $\sigma = 1$ . . . . .	91
2.39. Valores de $RRMSE$ (%) de los estimadores $\hat{\sigma}_R$ (usando los rangos muestrales y la constante $d_2$ ), $\hat{\sigma}_{R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\hat{\sigma}_S$ (usando la desviación típica muestral) y $\hat{\sigma}_P$ (usando la desviación típica ponderada) de $\sigma = 1$ . Datos seleccionados de la distribución Normal. . . . .	92
2.40. Valores de $RRMSE$ (%) de los estimadores $\hat{C}_{p,R}$ (usando los rangos muestrales y la constante $d_2$ ), $\hat{C}_{p,R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\hat{C}_{p,S}$ (usando la desviación típica muestral) y $\hat{C}_{p,P}$ (usando la desviación típica ponderada) de $C_p = 1$ . Datos seleccionados de la distribución Gamma con $\sigma = 1$ . . . . .	93
2.41. Valores de $RRMSE$ (%) de los estimadores $\hat{\sigma}_R$ (usando los rangos muestrales y la constante $d_2$ ), $\hat{\sigma}_{R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\hat{\sigma}_S$ (usando la desviación típica muestral) y $\hat{\sigma}_P$ (usando la desviación típica ponderada) de $\sigma = 1$ . Datos seleccionados de la distribución Gamma. . . . .	94
2.42. Valores de $RRMSE$ (%) de los estimadores $\hat{C}_{p,R}$ (usando los rangos muestrales y la constante $d_2$ ), $\hat{C}_{p,R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\hat{C}_{p,S}$ (usando la desviación típica muestral) y $\hat{C}_{p,P}$ (usando la desviación típica ponderada) de $C_p = 1$ . Datos seleccionados de la distribución Uniforme con $\sigma = 1$ . . . . .	95
2.43. Valores de $RRMSE$ (%) de los estimadores $\hat{\sigma}_R$ (usando los rangos muestrales y la constante $d_2$ ), $\hat{\sigma}_{R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\hat{\sigma}_S$ (usando la desviación típica muestral) y $\hat{\sigma}_P$ (usando la desviación típica ponderada) de $\sigma = 1$ . Datos seleccionados de la distribución Uniforme. . . . .	96



2.44. Valores de $RRMSE$ (%) de los estimadores $\widehat{C}_{pk.R}$ (usando los rangos muestrales y la constante $d_2$ ), $\widehat{C}_{pk.R2}$ (usando los rangos muestrales y las constantes $d_2$ and $d_2^*$ ), $\widehat{C}_{pk.S}$ (usando la desviación típica muestral) y $\widehat{C}_{pk.P}$ (usando la desviación típica ponderada) de $C_{pk} = 1$ . Datos seleccionados de la distribución Normal con $\sigma = 1$ en un proceso no centrado. . . . .	97
2.45. Valores de $RRMSE$ (%) de los estimadores $\widehat{\sigma}_R$ (usando los rangos muestrales y la constante $d_2$ ), $\widehat{\sigma}_{R2}$ (usando los rangos muestrales y las constantes $d_2$ and $d_2^*$ ), $\widehat{\sigma}_S$ (usando la desviación típica muestral) y $\widehat{\sigma}_P$ (usando la desviación típica ponderada) de $\sigma = 1$ . Datos seleccionados de la distribución Normal en un proceso no centrado. . . . .	98
2.46. Valores de $RRMSE$ (%) de los estimadores $\widehat{C}_{p.R}$ (usando los rangos muestrales y la constante $d_2$ ), $\widehat{C}_{p.R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\widehat{C}_{p.S}$ (usando la desviación típica muestral) y $\widehat{C}_{p.P}$ (usando la desviación típica ponderada) de $C_p = 1$ . Datos seleccionados de la distribución Gamma con $\sigma = 3$ . . . . .	99
2.47. Valores de $RRMSE$ (%) de los estimadores $\widehat{\sigma}_R$ (usando los rangos muestrales y la constante $d_2$ ), $\widehat{\sigma}_{R2}$ (usando los rangos muestrales y las constantes $d_2$ y $d_2^*$ ), $\widehat{\sigma}_S$ (usando la desviación típica muestral) y $\widehat{\sigma}_P$ (usando la desviación típica ponderada) de $\sigma = 3$ . Datos seleccionados de la distribución Gamma. . . . .	100
3.1. Sesgo teórico de la cuasidesviación típica muestral en un proceso con desviación típica $\sigma = 1$ y para distintos tamaños de muestra.	110
3.2. Gráficos QQ para cuantiles de la distribución probabilística $\chi_{n-1}^2$ y cuantiles del estadístico $(n - 1)\widehat{\sigma}^2/\sigma^2$ basados en muestras seleccionadas de poblaciones infinitas. Los tamaños muestrales oscilan entre $n = 5$ y $n = 250$ . . . . .	113
3.3. Gráficos QQ para cuantiles de la distribución probabilística $\chi_{n-1}^2$ y cuantiles del estadístico $(n - 1)\widehat{\sigma}^2/\sigma^2$ basados en muestras seleccionadas de poblaciones finitas con fracciones de muestreo ( $f = n/N$ ) entre el 1 % y el 50 %. El tamaño de la población es $N = 500$ , y los tamaños muestrales oscilan entre $n = 5$ y $n = 250$ .	114
3.4. Gráficos QQ para cuantiles de la distribución probabilística $\chi_{n-1}^2$ y cuantiles del estadístico $(n - 1)\widehat{\sigma}^2/\sigma^2$ basados en muestras seleccionadas de poblaciones finitas con fracciones de muestreo ( $f = n/N$ ) entre el 0.1 % y el 5 %. El tamaño de la población es $N = 10000$ , y los tamaños muestrales oscilan entre $n = 10$ y $n = 500$ . . . . .	115

# Índice de Tablas

2.1. Órdenes habituales de cuantiles usados en la construcción de intervalos de confianza con un nivel de significación $\alpha$ . . . . .	47
2.2. Valores del Sesgo Relativo ( $RB(\%)$ ) y Error Cuadrático Medio Relativo ( $RRMSE(\%)$ ) para los estimadores de $\sigma = 1$ (SD) y $C_p = 1$ (PCI), y cobertura empírica ( $Coverage(\%)$ ) de intervalos de confianza para $C_p = 1$ (PCI). Datos muestrales seleccionados de un distribución normal con $\mu = 10$ y $\sigma = 1$ . Para la población finita, $N = 500$ datos son seleccionados de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento. . . . .	54



# Capítulo 1

## Introducción

### 1.1. El control estadístico de calidad

Una definición tradicional de calidad de un producto o servicio es *la aptitud para su uso*. Es decir, se entiende que un producto o servicio tiene calidad si posee una o más características deseables que lo capacitan para satisfacer las necesidades de los usuarios que lo demandan. En este sentido, la calidad de los productos es un aspecto clave en la toma de decisiones del consumidor y, por ende, un aspecto clave para el éxito de la empresa.

Siguiendo la definición comentada, se puede clasificar la calidad como calidad de diseño y calidad de conformidad. La calidad de diseño hace referencia a los distintos niveles de calidad establecidos de forma intencionada y relacionados con el diseño del producto, más allá de su uso. Por otro lado, la calidad de conformidad se refiere a cómo de bien se ajusta el producto a las especificaciones del diseño previamente establecidas.

En esta investigación se asume una definición adicional de calidad utilizada por Montgomery (2009, p.6): la calidad es *inversamente proporcional a la variabilidad*. En relación a esta última definición, Bestelfield (2009, p.180) afirma que la variabilidad es inherente a cualquier proceso y, por tanto, los procesos de fabricación no suelen producir productos iguales. Las razones de este hecho son las distintas fuentes de variabilidad que pueden observarse en alguna de las fases de producción del producto, y que pueden ser atribuidas a diferentes causas. Besterfield (2009) enumeraba fuentes de equipo, material, ambiente y operario como principales responsables de dicha variabilidad. Es decir, la variación puede tener origen en: (i) el desgaste de las herramientas

o las vibraciones de las máquinas; (ii) el material empleado en la producción; (iii) la materia prima utilizada, puesto que puede contener diferencias entre unas y otras; (iv) el entorno; (v) condiciones ambientales, tal como el hecho de altas temperaturas, humedad; etc. Por último, también se puede mencionar a los propios operarios como parte de los causantes de la variabilidad de un proceso de producción, los cuales pueden cometer errores durante la fase de producción, diferentes turnos pueden producir productos diferentes, etc. De esta última definición de calidad se deduce que la mejora de la misma se identifica con la disminución de la variabilidad en los productos derivados de procesos de producción.

La calidad de los productos y servicios ha sido un aspecto que ha preocupado a las empresas y al ser humano, y en particular, desde el inicio de la producción. Hasta el siglo XIX, el aseguramiento de la calidad ha ido implícito en la fase de producción. A finales de dicho siglo, debido a la revolución industrial que trajo consigo el concepto de especialización laboral, se comienza a destinar operarios a inspeccionar la producción con el fin de evaluar y llevar un seguimiento de calidad de los productos. La década de 1920 es clave para el control estadístico de calidad ya que es cuando se introduce realmente el concepto y aparecen algunas de las grandes aportaciones de la época. Shewhart en 1924 desarrolla los gráficos de control, los cuales son las técnicas más utilizadas en el campo de control estadístico de calidad, y tiene como principal objetivo reducir la variabilidad del proceso, lo que hace que disminuya significativamente la proporción de artículos no conformes o defectuosos y, en consecuencia, mejora la productividad de la empresa (Besterfield 2014, Chakraborti et al. 2008, Chen 1997, Mitra 2008, Shewhart 1931). Dodge y Roming (1929) aportan la técnica de muestreo de aceptación como alternativa a inspeccionar todos los productos fabricados (Besterfield 2009 p. 361, Montgomery 2009 p.629).

En sus inicios, el control estadístico de la calidad no contó con demasiados seguidores, aunque en productos destinados a servicios militares tuvo una gran aceptación. Este hecho puede deberse en gran medida a que en la Primera Guerra Mundial se puso de manifiesto que el armamento alemán era más eficaz al utilizar unas tolerancias más estrictas, y ser dicho armamento más uniforme que el resto. Inglaterra, desde 1937, utiliza estas técnicas para la calidad de su producción. Estados Unidos creó la asociación “The American Society for Quality Control” en 1946. Esta asociación promovía el uso de técnicas de mejora de calidad y organizaba conferencias, publicaciones técnicas y programas de formación para la garantía de la calidad. Ese mismo año es fundada la Organización Internacional de Estándares (ISO) y el profesor Edward Deming de la Universidad de Columbia (Estados Unidos) es invitado a ayudar en la

reconstrucción de Japón en la que aplicó técnicas de control estadístico de calidad. Otros hechos de especial relevancia en este campo pueden encontrarse en la Tabla 1.1 de Montgomery (2009).

En la actualidad, la mayoría de las empresas utilizan controles de calidad debido a los beneficios que aporta. Instaurar un sistema de control de calidad permite el aseguramiento de la calidad en los productos ofrecidos al consumidor. Esto genera una buena reacción por parte de los consumidores y, como resultado, puede traducirse en un aumento de ventas. Asimismo, el control de calidad repercute positivamente en la productividad de la empresa, ya que estas técnicas además de mejorar el producto pueden hacer reducir costes de fabricación y producir ganancias reales a la empresa, compensando el gasto previo realizado en instaurar las técnicas de control de calidad.

El control de calidad puede definirse (véase Montgomery 2009 p.351, Mitra 2008, etc) como el conjunto de técnicas cuantitativas mediante las cuales se miden unas determinadas características de calidad de un producto, se comparan con unas especificaciones fijadas de antemano, y se toman acciones correctivas apropiadas cuando hay muchos elementos que no cumplen los requisitos o condiciones fijadas. Las características de calidad son por tanto una serie de propiedades que describen conjuntamente la aptitud de un producto o un servicio para su uso. Por simplicidad, en este trabajo hablaremos del control de calidad sobre un producto, aunque evidentemente esta técnica se puede aplicar en otros ámbitos tal como en el sector servicios. En adelante, se utilizan las siglas SQC, que corresponde a sus iniciales en inglés (*Statistical Quality Control*), para referirse al control estadístico de calidad, y tal como es habitual en la literatura existente sobre este tema.

Las características de calidad se pueden clasificar en variables o atributos en función de si son susceptibles de medición numérica o no, respectivamente. Destacamos que una característica de calidad clasificada como variable podría considerarse por conveniencia como atributo si se clasifican los productos como “apto” y “no apto” en función del valor de la característica de calidad. Por ejemplo, sea un proceso de fabricación que produce un determinado producto, y donde la característica de calidad es el peso de cada unidad producida. La empresa correspondiente podría considerar que el producto es apto si su peso supera una determinada cantidad o valor objetivo.

Estas características de calidad son evaluadas en relación a unas especificaciones. En este contexto, la característica de calidad tendrá un valor deseable que en control de calidad es llamado como valor objetivo. Algunos valores de la característica de calidad cercanos a este valor objetivo son también acep-

tables. De este modo, los límites de especificación vendrán determinados por el valor más alto aceptado, y el valor más bajo aceptado, denominados límite superior de especificación (LSE) y límite inferior de especificación (LIE), respectivamente. Estos límites son determinados de forma exógena al proceso. Pueden ser impuestos por agentes externos, tal como en el caso de la regulación del artículo, o por la propia empresa, como cuando ésta fija unos estándares de calidad. Por tanto, se dice que un producto tiene calidad en términos de conformidad cuando sus características de calidad presentan valores dentro de los límites de especificación. Cada falta de conformidad de una característica de calidad del producto se denomina defecto. Así, se distingue entre artículos no conformes, si el artículo es apto para desempeñar su función aunque presenta al menos una falta de conformidad, y defectuosos, si uno o más defectos presentes en el artículo impiden su uso.

Destacamos que los límites de especificación no tienen que ser necesariamente bilaterales. Por el contrario, existen multitud de situaciones donde se requieren límites de especificación unilaterales. Por ejemplo, una empresa que se dedica al embotellado de agua en botellas de 1 litro de capacidad podría requerir que éstas contengan al menos 0.95 litros de agua (LIE).

Una vez definida la característica de calidad de un producto, así como los límites de especificación, señalamos que el objetivo principal de este trabajo es el índice de capacidad de un proceso. Este índice compara la variabilidad de la característica de calidad con la amplitud de los límites de especificación, y es frecuentemente usado por las empresas para evaluar la capacidad de producir artículos conformes en sus procesos de producción.

En este trabajo denotaremos por  $X$  a la característica de calidad objeto de estudio y asumiremos que dicha variable  $X$  tiene una distribución Normal con media  $\mu$  y desviación típica  $\sigma$ , es decir,  $X \rightsquigarrow N(\mu, \sigma)$ , donde  $\mu$  denota la media del proceso y  $\sigma$  la desviación típica del proceso. Destacamos que es muy común el uso de esta hipótesis de normalidad en el contexto del SQC, tal como puede comprobarse en Montgomery (2009), Mitra (2008), Besterfield (2009), Chakraborti et al. (2008), Kotz y Lovelace (1998), etc. Existen varios trabajos y referencias que tratan sobre los efectos de la falta de Normalidad (véase, por ejemplo, Schilling y Nelson 1976, Chan et al. 1988 y Yourstone y Zimmer 1992). No obstante, en el campo del SQC se trabajan con muestras que pueden llegar a tener un tamaño considerable, y en este caso se podría hacer uso del Teorema Central del Límite (Montgomery 2009 p.86) para poder usar una aproximación válida a la distribución Normal en el caso de estadísticos como la media muestral, comúnmente utilizado en el contexto del SQC. Destacamos que, aunque el efecto de la falta de normalidad no es el tema principal del

trabajo, se analizará el impacto que tiene la ausencia de dicho supuesto en la estimación del índice de capacidad, así como en diferentes estimadores de la desviación típica del proceso. Además, se expondrán algunos casos de falta de normalidad que pueden presentarse en la práctica, pero que en realidad estas situaciones surgen por algún problema en una de las fases del proceso de producción, y por tanto la solución es realizar acciones correctivas sobre el proceso para que éste se encuentre en condiciones de normalidad y apto para la producción.

Puede observarse que bajo la hipótesis anteriormente comentada el proceso depende de los parámetros poblacionales  $\mu$  y  $\sigma$ , los cuales son utilizados en las distintas técnicas cuantitativas incluidas en el control estadístico de calidad, tal como en los gráficos de control, índices de capacidad, etc. En la práctica, tales parámetros suelen ser desconocidos y, en este caso, se recurre a la estimación de tales parámetros a partir de la información recogida de muestras del producto en estudio.

En control de calidad las variaciones se clasifican en: (i) variaciones debidas a causas fortuitas y (ii) variabilidad debida a causas atribuibles o asignables. El primer tipo de variaciones hacen referencia a la variabilidad natural presente en cualquier proceso, es decir, son consecuencia del efecto de pequeñas causas incontrolables acumuladas y que no suponen una gran variabilidad en el proceso. El segundo tipo de variación si provoca que el proceso de producción se encuentre en lo que se denomina, dentro del contexto del control estadístico de calidad, como fuera de control, es decir, son variaciones reconocibles e identificables y que suponen un aumento considerable de la variabilidad del proceso. En este caso, habría que tomar medidas inmediatas para corregir o disminuir esta variabilidad y llevar de nuevo al proceso a un estado denominado como bajo control.

Un concepto clave relacionado con la variabilidad natural de un proceso son los límites de tolerancia natural de un proceso. Estos límites representan los valores extremos de un intervalo en el que toma valores la variable del proceso,  $X$ . De forma general los límites de tolerancia natural se definen como:

- Límite Superior de Tolerancia Natural (LSTN) =  $\mu + c\sigma$
- Límite Inferior de Tolerancia Natural (LITN) =  $\mu - c\sigma$

donde  $c$  es un número arbitrario. En este trabajo se asume el método USA, el más común en el SQC, y en el cual se considera que  $c = 3$  (véase, por



ejemplo, Chakraborti et al. 2008, Chen 1997, Mitra 2008, Montgomery 2009, etc). De este modo, el intervalo de tolerancia natural tiene una amplitud de  $6\sigma$ . Suponiendo que la variable  $X$  sigue una distribución Normal este intervalo incluye el 99,73 % de las observaciones de la distribución, o dicho de otro modo, el 0,27 % está fuera del intervalo. En efecto, se tiene:

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = P(-3 < Z < 3) = 0,9973$$

donde  $Z$  denota la distribución Normal estándar.

Un proceso que presenta estas características se dice que tiene un nivel de rendimiento de calidad “tres sigma”. En el caso de que la característica de calidad  $X$  no siga la distribución Normal, el porcentaje de artículos no conformes será diferente al señalado con anterioridad.

Sin embargo, muchas empresas utilizan el criterio “Six-sigma”. Este concepto surge debido a que, si se consideran productos con múltiples características de calidad, un nivel de calidad tres sigmas puede no ser suficiente. Por ejemplo, si se piensa en un proceso de este nivel que fabrica un artículo complejo con, supongamos, 100 características de calidad, producirá solo 76,31 % de artículos conformes. Es decir, el 23,69 % de los artículos serán disconformes, cerca de uno cada cuatro de artículos producidos serán no conformes.

El concepto Six-sigma nace a finales de la década de 1980 y se ha ido desarrollando de tal forma que se identifican tres generaciones. La Generación I de Six-sigma tiene como objetivo reducir la variabilidad de los procesos al máximo y, para esto, se establecen unas tolerancias para el proceso de seis sigmas (seis veces la desviación típica del proceso) respecto a la media. De este modo, el proceso solo producirá 0.2 partes por millón (*ppm*) de artículos defectuosos:

$$P(\mu - 6\sigma < X < \mu + 6\sigma) = P(-6 < Z < 6) = 0,999999998$$

Además, este método asume desviaciones de 1.5 veces la desviación típica del proceso respecto a la media cuando el proceso se encuentra bajo control, es decir, debido a causas fortuitas, por lo que el objetivo real es alcanzar los 3,4 defectos por millón de unidades producidas (Breyfogle 2003, Pande et al. 2002). El mayor exponente de esta generación fue Motorola que mejoró considerablemente la proporción de artículos defectuosos producidos, lo que generó unos ahorros superiores a 940 millones de dólares en tres años (Hann

et al. 1999). La Generación II hace hincapié en vincular estos esfuerzos a los proyectos y actividades que mejoraron el rendimiento del negocio a través de la reducción de costes. Se incorpora la metodología DMAIC (Definir, Medir, Analizar, Mejorar y Controlar). La empresa General Electric es un referente en esta segunda generación ya que consiguió aumentar su margen operativo pasando del 14,4% al 18,4% durante los primeros cinco años de implantación del programa (Lucier et al. 2001). La Generación III añade el objetivo de crear valor a través de la organización y sus grupos de interés. Linderman et al. (2003, p.195) define Seis Sigma como un método sistemático y organizado para la mejora de los procesos estratégicos, que se basa en una metodología estadística y científica para lograr reducciones drásticas en las tasas de defectos definidas por los clientes. Debido a estos argumentos, las técnicas Six-sigma están recibiendo en la actualidad un mayor interés. Como se puede observar, las técnicas Six-sigma dependen del parámetro  $\sigma$ , el cual suele ser desconocido, tal como se argumentó con anterioridad. En este trabajo se analiza la estimación de la desviación típica de un proceso, por lo que las contribuciones realizadas también tienen un interés directo en el contexto del Six-sigma.

El control estadístico de calidad está basado en un conjunto de técnicas cuantitativas que se pueden clasificar en varios bloques según las funciones que desempeñan en la empresa. Esta investigación se centra en el control estadístico de procesos que incluye numerosas técnicas cuantitativas que tienen como objetivo controlar y evaluar un proceso de producción. Entre estas técnicas destacan los gráficos de control y el análisis de la capacidad del proceso.

### **1.1.1. Gráficos de control**

Los gráficos de control es una de las principales técnicas del control estadístico de procesos. El principal objetivo de los gráficos de control es controlar los parámetros asociados a la característica de calidad  $X$ . En general, las empresas están interesadas en controlar la variabilidad del proceso. En este sentido, los gráficos de control monitorizan el proceso productivo para detectar causas de variación asignables, lo que posibilita aplicar acciones correctoras lo antes posible, minimizando así las pérdidas generadas por la fabricación de artículos no conformes. Por tanto, los gráficos de control permiten identificar el estado del proceso.

Un paso previo a la elaboración de los gráficos de control es la selección de muestras aleatorias y periódicas del producto a inspeccionar. En cada artículo seleccionado, se mide la característica de calidad y, a continuación, se obtiene

para cada muestra el valor del estadístico correspondiente al parámetro del proceso que se desea controlar. De este modo se obtiene una serie temporal con el valor del estadístico para cada muestra seleccionada. Los gráficos de control consisten en representar gráficamente, sobre unos ejes cartesianos, la serie temporal del estadístico junto a la línea central y a los límites de control. El eje abscisas indica el número de orden cronológico de muestra y el eje de ordenadas, los valores del estadístico de cada muestra. El valor del estadístico para cada muestra estará representado por un punto, los cuales se van uniendo con los puntos de las muestras siguientes mediante líneas rectas entre cada dos puntos consecutivos. La línea central (LC) del gráfico de control es una línea horizontal que corresponde al valor esperado del estimador del parámetro poblacional objeto de estudio. El límite superior de control (LSC) y el límite inferior de control inferior (LIC) son dos líneas horizontales a una distancia predeterminada de la línea central que delimitan los valores que puede presentar el estadístico estando el proceso bajo control (Montgomery 2009 p.189, Besterfield 2003 p.105). Por ejemplo, en el caso de parámetros conocidos, usando el método USA y para el problema de controlar la media del proceso, los correspondientes límites de control están dados por:

$$LSC = \mu + 3 \frac{\sigma}{\sqrt{n}}$$

$$LIC = \mu - 3 \frac{\sigma}{\sqrt{n}}$$

donde  $n$  denota el tamaño de cada muestra seleccionada para la realización del gráfico de control. El estadístico utilizado en este gráfico de control para controlar la media del proceso es la media muestral, donde es conocido que su valor esperado es la media poblacional,  $\mu$ , y por tanto este valor determina la LC del gráfico de control. Los límites de control se fijan a una distancia de 3 veces la desviación típica del estadístico, la cual viene determinada por  $\sigma/\sqrt{n}$ . Destacamos que el valor 3 es el utilizado por defecto por el método USA.

Se dice que un proceso está bajo control si todos los puntos representados en el gráfico de control se sitúan en la zona delimitada por los límites de control y se distribuyen de forma aleatoria, sin rachas ascendentes o descendentes de puntos, tal como puede verse en la Figura 1.1. Cuando el proceso se encuentra en esta situación, el proceso es estable y predecible, es decir, la variabilidad dentro del mismo es debido exclusivamente a causas fortuitas y la uniformidad entre artículos es máxima. Besterfield (2009 p.130) enumera algunos de los beneficios que se presentan cuando el proceso se encuentra en este estado. Por ejemplo, la posibilidad de disminuir al mínimo el coste por inspección, puesto

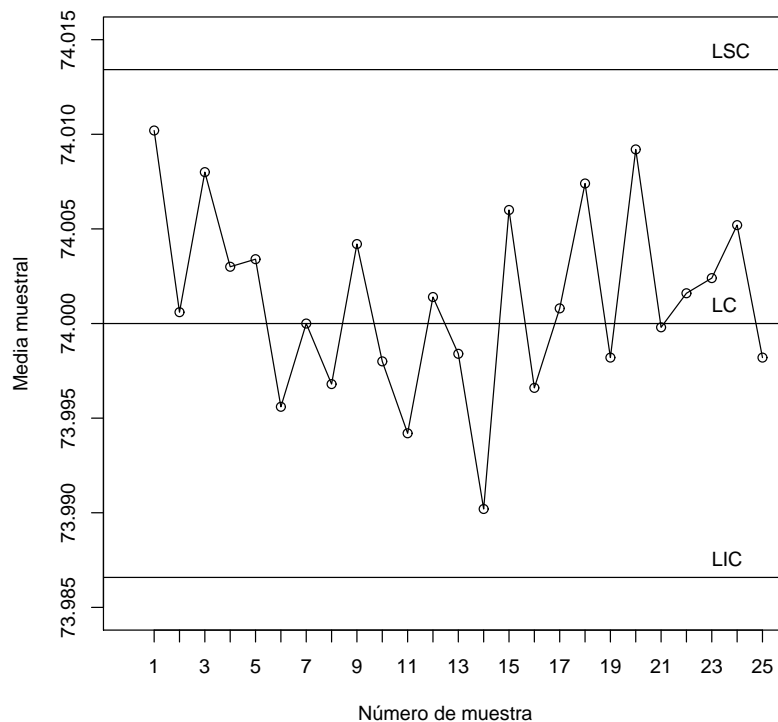


Figura 1.1: Gráfico de Control para la media de un proceso que se encuentra bajo control. Datos reales tomados de la base de datos “pistonrings” incluidos en la librería “qcc” (Scrucca 2004) del software estadístico R.

que se necesitan menos muestras para evaluar la calidad. Por el contrario, si algún valor del estadístico se sitúa por encima del  $LSC$  o por debajo del  $LIC$ , el proceso estará en un estado de fuera de control. Además, aunque todos los valores del estadístico estén situados en la zona delimitada por ambos límites, si éstos no se distribuyen de manera aleatoria, el proceso también está fuera de control. Pueden consultarse algunas reglas útiles para determinar si el proceso está fuera de control en Montgomery (2009, p.197), Besterfield (2003, p.131) o Hurwitz y Mathur (1992). Algunas reglas muy comunes son la presencia de rachas ascendentes o descendentes de puntos, cuatro de cinco puntos consecutivos situados a una distancia mayor de  $\sigma$  de la  $LC$ , una serie de ocho puntos consecutivos a un lado de la  $LC$ , catorce puntos alternando arriba y abajo o un patrón inusual en los datos.

Los gráficos de control han sido utilizados por multitud de empresas ya que son especialmente útiles gracias a los beneficios que aportan. Montgomery (2009, p.189) enumera cinco razones por las que su uso es tan común:

1. Los gráficos de control mejoran la productividad de la empresa, debido a que reducen considerablemente la proporción de artículos defectuosos. Como se ha comentado, los gráficos de control tienen como objetivo principal reducir la variabilidad de los productos producidos, lo que produce que una mayor cantidad de artículos producidos sean conformes a las especificaciones. En consecuencia, un menor número de artículos producidos son desechados, por lo que se reducen los costes en la producción y aumenta la capacidad de producción, es decir, el proceso es capaz de producir una mayor cantidad de artículos conformes en un mismo intervalo de tiempo. Todo esto contribuye a tener un proceso de producción más eficaz y rentable.
2. Los gráficos de control son efectivos en la prevención de defectos. Un cambio de estado en un proceso que se encuentra bajo control es rápidamente detectado por los gráficos de control. Esto provoca una rápida intervención en el proceso y, por tanto, reduce la cantidad de artículos defectuosos que produciría el proceso en estado de fuera de control. Así, los gráficos de control ayudan a la empresa a ser más eficiente, ya que permite una producción correcta desde el primer momento.
3. Los gráficos de control reducen los costes derivados de una sobreintervención en el proceso. En otras palabras, los gráficos de control son capaces de diferenciar variabilidad debida a causas fortuitas y la variabilidad debida a causas asignables, por lo que es posible conocer cuándo puede mejorar el proceso como consecuencia de una acción correctiva. Por ende, solo se intervendrá en el proceso en este último caso.
4. Los gráficos de control aportan información para el análisis de la causa de variabilidad por parte del ingeniero u operario. Una vez que se detecta que el proceso está fuera de control, la siguiente fase es analizar la causa que ha provocado este estado. Los gráficos de control ayudan a identificar la causa responsable y, de este modo, a la elección de la acción correctiva a llevar a cabo que devuelva el proceso a un estado de bajo control.
5. Los gráficos de control aportan información sobre la capacidad del proceso. En este sentido, aportan información sobre los parámetros del proceso y su estabilidad en el tiempo.

Este último punto pone de manifiesto la relación existente entre los gráficos de control y el análisis de la capacidad de un proceso, que es el principal tema de estudio en este trabajo. De hecho, no parece razonable evaluar un proceso que se encuentre en situación de fuera de control, ya que existen en este proceso causas asignables. En este contexto, Besterfield (2009, p.129) afirma que la verdadera capacidad del proceso se logra cuando el proceso está bajo control. En conclusión, el análisis de la capacidad solo debe realizarse sobre procesos que han alcanzado su auténtica capacidad, que se consigue cuando el proceso está bajo control, es decir, la variabilidad solo es atribuible a causas fortuitas.

## 1.2. Análisis de la capacidad de un proceso

Recordamos que el análisis de la capacidad de un proceso sirve para evaluar la aptitud de dicho proceso para producir artículos dentro de las especificaciones fijadas de antemano. Un análisis de capacidad de un proceso puede estar basado en muchas técnicas (histogramas, gráficos de normalidad, etc), pero el indicador más común para su medición es el índice de capacidad de un proceso (en adelante *PCI*, “Process Capability Index”), y también llamado como el ratio de la capacidad de un proceso. El PCI puede ser de gran importancia para muchas empresas, y ha sido ampliamente estudiado por esta razón (Anis 2008, Besseris 2014, Bissell 2013, Boyles 1991, Chan et al. 1988, Chen y Ding 2001, Chen et al. 2001, Chen et al. 2003, English y Taylor 1993, Kane 1986, Kotz y Johnson 2002, Kotz y Lovelace 1998, Kushler y Hurley 1992, Luceño 1996, Montgomery 2009 p.351, Pearn et al. 1992, Porter y Oaklnad 1991, Rodriguez 1992, Somerville y Montgomery 1996, Spiring et al. 2003, Yeh y Bhattacharya 1998, etc.). Por ejemplo, el PCI puede determinar si el proceso es válido en la producción de un artículo bajo ciertos requisitos de calidad, y un diagnóstico previo es un importante paso para reducir costes y mejorar la productividad. Por tanto, el PCI evalúa la capacidad de un proceso para producir artículos dentro de los límites de especificación prefijados, es decir, la capacidad de producir productos aceptables. Es conveniente destacar que el PCI, como el resto de técnicas correspondientes al análisis de la capacidad de un proceso, solo es aplicable a procesos que están en un estado de control estadístico, es decir, en procesos en situación de bajo control (véase, por ejemplo, la Sección 1.1.1, Montgomery 2009 p.181, Evans y Lindsay 1999 p.321, Jensen et al. 2006).

Sea  $X$  la característica de calidad asociada a un proceso que opera con una media  $\mu$  y una desviación típica  $\sigma$ . El *PCI* se define como el ratio de la amplitud de los límites de especificación y la amplitud de los límites de

tolerancia natural del proceso, es decir,

$$C_p = \frac{LSE - LIE}{6\sigma} \quad (1.1)$$

donde  $LSE$  y  $LIE$  son, respectivamente, el límite superior e inferior de especificación, ya definidos con anterioridad, y  $6\sigma$  representa la variabilidad natural del proceso, y la cual viene determinada por la diferencia de los límites superior e inferior de tolerancia natural, es decir,

$$LSTN - LISN = \mu + 3\sigma - (\mu - 3\sigma) = 6\sigma.$$

Por tanto, el  $PCI$  compara la variabilidad natural del proceso con el rango de valores permitidos en el proceso, que es definido por los límites de especificación.

El  $PCI$  describe cómo de bien el proceso produce artículos aceptables, es decir, la proporción de artículos que están dentro de los límites de especificación. El proceso puede clasificarse como no apto si el  $PCI$  es menor que 1. En este caso la amplitud de los límites de especificación es menor que la variabilidad natural del proceso, y esto implica que el proceso produce una proporción excesiva de artículos no conformes (véase la Figura 1.2 (a)). En esta situación se requieren acciones correctivas sobre el proceso. Por ejemplo, si hay margen en los requisitos de calidad, la acción correctiva más fácil de aplicar para reducir la proporción de artículos no conformes es incrementar el rango de los límites de especificación. De lo contrario, se deben introducir algunas mejoras en el proceso para reducir la variabilidad y la proporción de elementos no conformes.

El proceso es adecuado para producir artículos dentro de los límites de especificación si el  $PCI$  es mayor que 1. La proporción de artículos no conformes debe ser pequeña en esta situación, y disminuye a medida que el valor de  $PCI$  aumenta (véase la Figura 1.2 (b)). Además, un cambio pequeño de los parámetros asociados con el proceso tendrá un impacto menor en la proporción de elementos no conformes. En este sentido, Montgomery (1985) señala que las empresas deberían considerar un índice de capacidad mínimo fijado en  $4/3$  para procesos que llevan un tiempo considerable en funcionamiento, y un valor de  $5/3$  para nuevos procesos de producción.

El proceso es marginalmente adecuado para producir elementos dentro de los límites de especificación si el  $PCI$  es igual a 1 (Kotz y Johnson 1993, Chakraborti et al. 2008 y Chen 1997). En esta situación el proceso es muy sensible

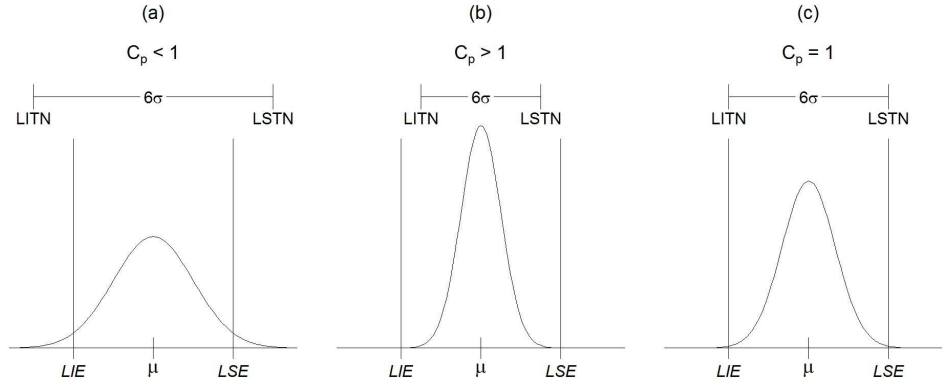


Figura 1.2: Relación entre límites de especificación y límites de tolerancia natural según el *PCI*.

a cualquier cambio en los parámetros asociados con éste, puesto que este problema producirá una aumento significativo en la proporción de elementos no conformes, tal como puede observarse en la Figura 1.2 (c).

En la ecuación (4.2) se observa que el índice  $C_p$  utiliza límites de especificación bilaterales, que es la situación más común. Sin embargo, los índices de capacidad unilaterales pueden ser fácilmente definidos (Montgomery 2009 p.352):

- Usando sólo el límite superior de especificación:  $C_{pu} = \frac{LSE - \mu}{3\sigma}$
- Usando sólo el límite inferior de especificación:  $C_{pl} = \frac{\mu - LIE}{3\sigma}$

Por simplicidad y por ser la situación más común, en este trabajo se considera el caso de límites de especificación bilaterales, pero la extensión de esta investigación al caso de límites de especificación unilaterales es simple, y se esperan también resultados similares.

El *PCI* definido por la expresión (4.2) asume que el punto medio de los límites de especificación es igual a la media del proceso  $\mu$ . En esta situación se dice que el proceso está centrado. En caso contrario se dice que el proceso es no centrado y el *PCI* debe tener en consideración la posición de la media del proceso respecto a los límites de especificación. Por esta razón, el índice de capacidad de un proceso para un proceso no centrado ( $PCI_k$ ) se define como



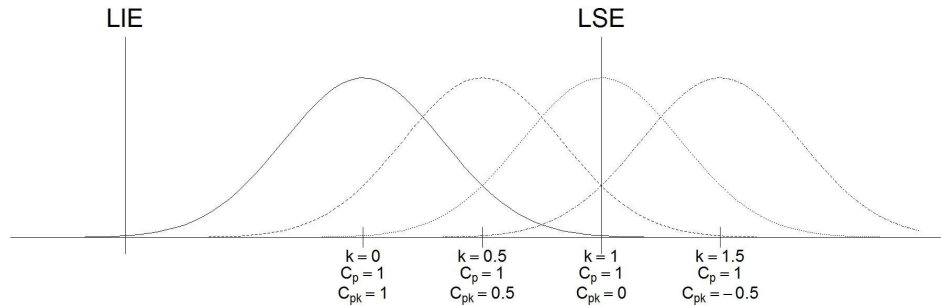


Figura 1.3: Diferencias entre  $C_p$  y  $C_{pk}$  en función del valor de  $k$  y para unos determinados límites de especificación.

$$C_{pk} = (1 - k)C_p$$

donde

$$k = \frac{|\mu_0 - \mu|}{(LSE - LIE)/2},$$

y

$$\mu_0 = \frac{LIE + LSE}{2}$$

es el punto medio de los límites de especificación. En función del valor de  $k$  pueden darse hasta tres situaciones diferentes representadas gráficamente en la Figura 1.3. El valor  $k = 0$  indica que el proceso productivo está centrado, es decir,  $\mu_0 = \mu$  y  $C_{pk} = C_p$ . Si  $0 < k < 1$ , la distancia entre  $\mu_0$  y  $\mu$  es menor que la mitad del intervalo definido por los límites de especificación, y  $C_{pk} < C_p$ . Si  $k \geq 1$ , la distancia entre  $\mu_0$  y  $\mu$  es mayor que la mitad del intervalo definido por los límites de especificación, y esto implica que la proporción de artículos no conformes es demasiado elevada.

### 1.3. Objetivos

En secciones anteriores se ha descrito de forma muy breve el concepto de control estadístico de calidad y se han definido dos de las herramientas más

comunes del control estadístico de procesos: los gráficos de control y el análisis de la capacidad de un proceso. Una vez aclarados estos conceptos, puede decirse que el objetivo principal de este trabajo es el análisis de capacidad de un proceso bajo escenarios que no han sido analizados previamente en la literatura existente. Para ello se describen a continuación una serie de objetivos intermedios que se persiguen en este trabajo:

**O1. Evaluar el comportamiento empírico del estimador de la desviación típica del proceso y su efecto en la estimación del *PCI*.**

Como se comentó anteriormente, en el SQC los procesos basados en normalidad dependen de los parámetros poblacionales  $\mu$  y  $\sigma$ . En la práctica estos parámetros son desconocidos, por lo que se recurre a la estimación puntual de los mismos. La estimación puntual del parámetro  $\sigma$  es un tema de especial interés en este trabajo, puesto que la capacidad de un proceso está directamente relacionada con la variabilidad de dicho proceso. En particular, el *PCI* compara los límites de especificación con la variabilidad natural de un proceso (véase la ecuación (4.2), y el cual depende directamente de la desviación típica del proceso. Para estimar el *PCI* definido por la ecuación (4.2) tan sólo hay que sustituir  $\sigma$  en esta expresión por su correspondiente estimador. Este hecho revela la importancia de la estimación de  $\sigma$  en el análisis de la capacidad de un proceso basándonos en muestras aleatorias. Asumiendo la selección de una sola muestra para la estimación de  $\sigma$ , el estimador más común viene dado por la desviación típica muestral. Este estimador no tiene la propiedad de insesgadez, por lo que usualmente se recurre a la versión insesgada. El objetivo O1 de esta investigación consiste en evaluar empíricamente ambos estimadores de  $\sigma$ , con el fin de analizar, entre otros aspectos, el efecto de la corrección para la insesgadez. Además se evaluará el comportamiento de los estimadores del *PCI* basados en ambos estimadores de  $\sigma$ . Este estudio se realizará mediante estimación puntual y por intervalos de confianza. Para alcanzar dicho objetivo se ha llevado a cabo un estudio empírico del comportamiento de los estimadores en diferentes escenarios, y se han comentado las principales conclusiones (ver la Sección 2.2.3).

**O2. Evaluar el estimador del *PCI* en procesos no centrados.**

Una vez analizado el comportamiento de los estimadores puntuales de  $\sigma$  y el *PCI*, y la estimación mediante intervalos de confianza en procesos centrados, parece razonable que el siguiente objetivo sea evaluar tales estimadores en procesos no centrados. Además, resulta conveniente analizar si influye en la calidad de la estimación el grado de descentralización

del proceso, es decir, si es relevante la distancia a la que se encuentre la media del proceso del punto medio de los límites de especificación. En este sentido, se han creado varios escenarios con distintos valores de  $k$ . Se parte de un primer escenario en el que  $k = 0$ , es decir, el proceso es centrado. Este escenario sirve de referencia para poder comparar el comportamiento de los estimadores en procesos no centrados. Un segundo escenario considera  $k = 0.25$ , es decir, la media del proceso está desplazada respecto al punto medio de los límites de especificación una octava parte de la amplitud del intervalo formado por los mismos. En el tercer escenario se toma  $k = 0.5$ , es decir, la media del proceso se encuentra en un punto intermedio entre el punto medio de los límites de especificación y uno de tales límites de especificación. Utilizando estos escenarios y asumiendo una sola muestra, en la Sección 2.2.3 se realiza un estudio del comportamiento empírico del estimador insesgado de  $\sigma$ , y el estimador puntual y por intervalos de confianza del *PCI* basado en este estimador.

**O3. Analizar el efecto en la estimación de la desviación típica y del *PCI* cuando el proceso está fuera de control.**

La estimación del *PCI* se debe realizar cuando se satisfacen unos supuestos de partida (ver Sección 2.2.1). Por ejemplo, uno de ellos es que el proceso debe encontrarse en una situación de bajo control cuando se lleva a cabo el análisis de capacidad, es decir, para determinar el *PCI*. La idea que subyace es que solo cuando el proceso se encuentra en este estado está rindiendo a su verdadera capacidad. El objetivo O3 es analizar el efecto que tiene el incumplimiento de este supuesto básico en la estimación del *PCI*. Para abordar este objetivo se ha llevado a cabo una simulación Monte Carlo en la que se comparan los resultados de la estimación del índice de capacidad cuando se consideran solo las muestras que se obtienen de un proceso bajo control, frente a la estimación del índice de capacidad cuando se consideran todas las muestras, es decir, tanto las que proceden de situaciones en las que el proceso está bajo control, como en escenarios de fuera de control.

**O4. Analizar el comportamiento empírico del estimador de la desviación típica y *PCI* en poblaciones infinitas y en poblaciones finitas.**

Otro supuesto de partida (ver también Sección 2.2.1) es la independencia e idéntica distribución de la variable asociada a los valores muestrales. Este supuesto básico se presenta en el caso de poblaciones infinitas. Sin embargo, las empresas tienen una producción finita y las muestras suelen obtenerse de lotes finitos de productos, por lo que resulta interesante

analizar el impacto que tiene el incumplimiento de este supuesto básico de partida en la estimación del *PCI*. Cabe destacar que otras técnicas cuantitativas utilizadas en el SQC asumen lotes de producción finitos, como por ejemplo el muestreo de aceptación. Para lograr este objetivo, se tuvo en cuenta escenarios con poblaciones infinitas y finitas en el estudio empírico llevado a cabo en la Sección 2.2.

**O5. Evaluar el comportamiento empírico del estimador de la desviación típica y del *PCI* en poblaciones “aproximadamente” infinitas (con una fracción de muestreo muy pequeña).**

En relación con el supuesto de partida comentado en O4, una situación interesante ocurre cuando la muestra es relativamente pequeña en comparación con el tamaño de la población, es decir, cuando se produce una aproximación a una población infinita. El objetivo O5 consiste en evaluar los diferentes estimadores en este caso. Para alcanzar este objetivo y analizar mejor las diferencias con respecto otros ambientes, se ha ampliado el estudio donde se considera población finita. Por un lado se crea un escenario en el que la muestra se selecciona de poblaciones finitas con fracciones de muestreo relativamente altas. Esto implica una diferencia significativa entre esta situación y la que considera población infinita. Por otro lado, se crea un escenario en el que la muestra se selecciona de poblaciones finitas con fracciones de muestreo próximas a cero. De esta manera se obtiene un caso similar al que considera la población infinita.

**O6. Analizar mediante gráficos QQ la distribución de la cuasivarianza muestral en los diferentes escenarios.**

Como es sabido, el estadístico

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2$$

cuando los elementos muestrales son independientes e idénticamente distribuidos (i.i.d.) y proceden de una distribución Normal (Herrerías et al. 2012 p.66). Este estadístico está estrechamente relacionado con la estimación por intervalos de confianza del índice de capacidad, y el incumplimiento de esta distribución es de gran importancia. El objetivo O6 consiste en analizar la distribución de este estadístico en diferentes situaciones. Para alcanzar este objetivo se han elaborado gráficos QQ (Beirlant et al. 2006 p.3, Wilk y Gnanadesikan 1968) basados en muestras seleccionadas de poblaciones infinitas, poblaciones finitas con fracciones de muestreo relativamente altas, y poblaciones finitas donde

la fracción muestral es proxima a cero, es decir, aproximación a una población infinita.

**O7. Estudiar empíricamente del estimador de la desviación típica y del  $PCI$  en poblaciones que no presentan normalidad.**

Como se expuso anteriormente, en el SQC es habitual asumir que la característica de calidad sigue una distribución Normal. De hecho, este es otro supuesto básico de partida en la estimación del  $PCI$ . El objetivo O6 consiste en evaluar los diferentes estimadores ante el incumplimiento de este supuesto. Para lograr este objetivo se ha incluido en el estudio de simulación llevado a cabo en la Sección 2.2 dos escenarios alternativos en los que los datos proceden de una distribución Gamma y de una distribución Uniforme. La elección de la distribución Gamma es debido a que se trata de una distribución asimétrica a la derecha, a diferencia de la distribución Normal donde es conocida su simetría, y con la presencia de datos anómalos en la parte derecha de su distribución. Mientras que la distribución Uniforme ha sido elegida por representar un caso extremo a la distribución Normal.

**O8. Evaluar el comportamiento empírico del estimador de la desviación típica y del  $PCI$  en una población basada en datos reales.**

Adicionalmente, un punto de interés es evaluar los estimadores de los índices de capacidad en un entorno real. Para ello se ha utilizado el conjunto de datos reales “pistonrings” incluido en la librería “qcc” (Montgomery 2009 p.251, Scrucca 2004) del software estadístico R, y se ha llevado a cabo un estudio de simulación donde se evalúa el comportamiento empírico de los estimadores en estudio en este trabajo. Además, con el objetivo de analizar las diferencias entre poblaciones finitas e infinitas en este conjunto real de datos, se seleccionan muestras aleatorias simples con y sin reemplazamiento. Destacamos que para muestras sin reemplazamiento no se verificará la hipótesis básica de independencia. Por otro lado, se han seleccionado muestras mediante muestreo aleatorio simple con reemplazamiento, lo cual garantiza la independencia entre los elementos muestrales, y la situación será similar al caso de una población infinita. Asimismo, en cada caso se han comparado los resultados de las estimaciones del índice de capacidad cuando se considera que las muestras proceden de situaciones en las que el proceso está bajo control frente a situaciones en las que el proceso está fuera de control. En resumen, el objetivo O8 consiste en evaluar el comportamiento empírico de los estimadores en un entorno real en el que se dan las características de los escenarios estudiados en los objetivos anteriores.

**O9. Estudiar comparativamente los principales estimadores de la desviación típica y el  $PCI$  usando  $m$  muestras.**

Una opción que también se contempla en la literatura del SQC y es comúnmente usada y aceptada en la práctica es utilizar la información recogida en  $m$  muestras, donde  $m > 1$ , en lugar del caso de una única muestra usado hasta el momento. Por ejemplo, los gráficos de control con parámetro desconocido hacen uso de  $m$  muestras para la estimación de sus parámetros. En la Sección 2.3.2 se presentan los principales estimadores de la desviación típica y  $PCI$  asumiendo  $m$  muestras, y se lleva a cabo un estudio de simulación en el que se compara el comportamiento empírico de dichos estimadores. Además, se ha incluido un escenario en el que el proceso es no centrado.

**O10. Estudiar comparativamente los principales estimadores de la desviación típica y el  $PCI$  usando  $m$  muestras en poblaciones que no presentan normalidad.**

Para alcanzar este objetivo se han usado dos situaciones alternativas al caso principal en el que los datos muestrales proceden de una distribución Normal: datos procedentes de las distribuciones Gamma y Uniforme. Este estudio permite llevar a cabo un completo análisis comparativo entre los principales estimadores en diferentes situaciones, y obtener una serie de conclusiones sobre el comportamiento de estos estimadores en distintos ambientes (ver Sección 2.3.3).

**O11. Proponer nuevas expresiones para la constante  $c_4$ .**

Algunos de los estimadores comúnmente utilizados presentan la propiedad de insesgadez, tal como la media muestral o la cuasivarianza muestral. Por el contrario, como ya se ha mencionado, la cuasidesviación típica muestral no presenta esta propiedad, por lo que es necesario realizar una corrección en este estimador para obtener la insesgadez. Esta transformación consiste en dividir la cuasidesviación típica muestral por una constante, a la cual se le denomina  $c_4[\cdot]$  en el contexto del SQC. Esta constante asume su valor en función del tamaño muestral y se determina mediante una expresión que incluye la función Gamma (Herrerías et al. 2012 p.14). Sin embargo, la función Gamma tiende a infinito a medida que aumenta el tamaño muestral, por lo que no resulta posible el cálculo de la constante  $c_4[\cdot]$  en esta situación y a través de su propia definición. A modo de ejemplo, a partir de tamaños muestrales superiores a 343, el programa estadístico R (R Core Team 2016) proporciona valores infinitos para esta constante. El objetivo O11 de esta investigación es proponer

nuevas expresiones para la constante  $c_4$  que posibilite su cálculo para tamaños muestrales elevados de forma directa.

**O12. Elaborar una librería en el programa estadístico R para el análisis de la capacidad de un proceso.**

El programa estadístico R (R Core Team 2016) es un entorno interactivo para el análisis estadístico y gráfico que, gracias a las ventajas que presenta, se está utilizando en trabajos de investigación de numerosas áreas y, en especial, las Ciencias Sociales. Se compone de un sistema base de funciones que realizan multitud de operaciones estadísticas, y de paquetes o librerías adicionales, es decir, una librería es un conjunto de funciones relacionadas con un tema concreto, y que permite de este modo extender la funcionalidad del software. R es un programa gratuito en el que se puede contribuir elaborando nuevas librerías, las cuales son evaluadas por responsables del mantenimiento de R, y si es el caso, las nuevas librerías son añadidas en R. De este modo, R se actualiza de forma constante e incorpora las técnicas de investigación más actuales. Pese a que R cuenta con varias librerías relacionadas con el SQC, como “IQCC” (Recchia et al. 2010) o “qcc” (Scrucca 2004), no tiene ninguna librería que se centre exhaustivamente en el análisis de capacidad de un proceso. Ante esta carencia, estamos desarrollando la librería “Capability” que se ocupa de esta temática e incorpora recientes metodologías de estimación puntual y por intervalos de confianza para el índice de capacidad.

## Capítulo 2

# Análisis del índice de capacidad en distintos escenarios

### 2.1. Introducción

Los índices de capacidad utilizados para el análisis de la capacidad de un proceso se basan en la desviación típica de dicho proceso. Como se expuso en el capítulo anterior, el índice de capacidad potencial de un proceso (*PCI*) es el ratio o cociente entre la amplitud de los límites de especificación (*LIE*; *LSE*) y la variabilidad natural del proceso. Los límites de especificación son definidos de manera exógena al proceso productivo. Por su parte, la variabilidad natural del proceso es común definirla como seis veces la desviación típica del proceso ( $6\sigma$ ). En definitiva, el *PCI* suele obtenerse como

$$C_p = \frac{LSE - LIE}{6\sigma} \quad (2.1)$$

En la práctica, la desviación típica del proceso es un parámetro desconocido que debe ser estimado. Principalmente existen dos metodologías para la estimación de este parámetro. Por una parte, seleccionar una muestra de la producción y estimar el valor del parámetro con la información obtenida a partir de esta única muestra. Por otra parte, se pueden utilizar  $m$  muestras, donde  $m > 1$ , y las cuales podrían ser las empleadas en la elaboración de los gráficos de control. En general, las muestras comentadas se seleccionan mediante muestreo aleatorio simple sin reemplazamiento (Cochran 1977, Särndal et al. 1992). Además, las muestras han de ser seleccionadas cuando se veri-



fican una serie de supuestos, entre los que se encuentra, que el proceso esté bajo control (véase la Sección 1.1.1). La estimación del índice de capacidad definido en la ecuación (2.1) se obtiene sustituyendo en dicha ecuación  $\sigma$  por el correspondiente estimador.

Por una parte, en este capítulo se lleva a cabo un estudio empírico basado en el principal estimador de la desviación típica, y asumiendo el escenario de una única muestra. Se evaluará el impacto del incumplimiento de cada uno de los supuestos básicos comúnmente utilizados para la estimación del índice de capacidad de un proceso. Estos supuestos básicos están descritos en la Sección 2.2.1. En este sentido, se realizará un completo estudio de simulación Monte Carlo en el que se diseñarán los escenarios oportunos para alcanzar dicho objetivo. Por último, se expondrán las principales conclusiones obtenidas de dicho estudio.

A modo de extensión, en la Sección 2.3 se llevará a cabo un análisis comparativo del comportamiento empírico de los principales estimadores de la desviación típica y del índice de capacidad cuando se seleccionan  $m$  muestras para la estimación de parámetros. Esta comparación se realizará asumiendo que se cumplen todos los supuestos de partida. Además, se evaluará el comportamiento de estos estimadores bajo el incumplimiento del supuesto de normalidad. Para ello, se analizarán poblaciones que siguen distribuciones probabilísticas Gamma y Uniforme.

## 2.2. Caso de una única muestra

La selección de una muestra para la estimación de parámetros relacionados con el índice de capacidad puede suponer un menor coste que la selección de múltiples muestras, por lo que es probable que una organización se decante por esta opción si tiene la certeza de estar operando con un proceso productivo en situación de bajo control, y la muestra sea lo suficientemente representativa. Se han propuesto numerosos métodos para la estimación de la desviación típica del proceso  $\sigma$ . Algunos de los más conocidos y usados son los basados en la desviación típica muestral y los rangos muestrales (Chakraborti 2008, Chen 1997, Duncan 1986, Jones et al. 2001, Luko 1996, Montgomery 2009 pp.118 y 253, Ott 1975, Vanderman 1999, Wheeler 1995, Woodall y Montgomery 2000, Muñoz et al. 2016). En esta sección se asume la selección de una única muestra para la estimación del PCI, y se considera el estimador puntual basado en la desviación típica muestral. Además, se analiza el comportamiento del intervalo de confianza tradicional para el PCI basado en este estimador (Kotz

y Johnson 1993 p.43). Una contribución añadida que se lleva a cabo es analizar el efecto en la estimación del PCI y su correspondiente intervalo de confianza considerando diferentes estimadores de la desviación típica del proceso. En particular, se considera el estimador sesgado y el estimador insesgado de la desviación típica del proceso basados, ambos en la desviación típica muestral. El objetivo principal de este último estudio es analizar los sesgos de ambos estimadores bajo distintos tamaños muestrales, y comprobar si el sesgo del estimador sesgado toma valores razonables en el caso de muestras pequeñas.

El objetivo principal de esta sección es analizar el PCI cuando algunos de los supuestos básicos no son completamente satisfechos. En particular, el estimador del PCI y/o el correspondiente intervalo de confianza asumen que: (i) los datos proceden de una distribución Normal; (ii) las muestras son extraídas de poblaciones infinitas; y (iii) el proceso está bajo control. Estos supuestos se describen en la Sección 2.2.1. Destacar que el incumplimiento de estos supuestos es una situación que puede plantearse en la realidad. Por ejemplo, las distribuciones probabilísticas asociadas a los estadísticos son válidas cuando las variables son independientes e idénticamente distribuidas. Sin embargo, pueden plantearse situaciones donde este supuesto no se cumple. Por ejemplo, la producción de numerosas organizaciones se almacena en lotes finitos, de los cuales se extraerán las muestras para la aplicación de diferentes técnicas de control estadístico de calidad. El muestreo de aceptación es un ejemplo del contexto del SQC donde las muestras se seleccionan de poblaciones finitas (Montgomery 2009 p.629, Aslam et al. 2015, Rao et al.2016, Nezhad y Niaki 2013, etc). Si las muestras son seleccionadas de poblaciones finitas sin reemplazamiento, el intervalo de confianza usual para el PCI, que está basado en poblaciones infinitas, debería tener un peor comportamiento, especialmente para fracciones de muestreo elevadas. Una contribución de esta sección es analizar numéricamente el intervalo de confianza tradicional para el PCI, el cual asume poblaciones infinitas. Sin embargo, se considera una situación común en la práctica (poblaciones finitas) para investigar el efecto de el incumplimiento de este supuesto. Finalmente, también se comparan los resultados derivados de procesos centrados y no centrados. Una descripción detallada de las contribuciones de esta sección puede verse en la Sección 2.2.3.

En esta Sección 2.2 asume una única muestra con un tamaño  $n$ , donde  $x_1, x_2, \dots, x_n$  denotan los valores muestrales de la característica de calidad  $X$ . La media muestral se define como

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (2.2)$$

mientras la desviación típica muestral se define como

$$S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2}. \quad (2.3)$$

### 2.2.1. Supuestos básicos

La estimación del PCI está basada en los siguientes supuestos básicos:

- A1. La característica de calidad  $X$  sigue una distribución Normal, es decir,  $X \rightsquigarrow N(\mu, \sigma)$ , donde  $\mu$  es la media con la que opera el proceso y  $\sigma$  es su desviación típica.
- A2. Las variables asociadas a los valores muestrales  $x_1, x_2, \dots, x_n$  son independientes e idénticamente distribuidas.
- A3. El proceso está bajo control (ver Montgomery 2009 p.181, Evans y Lindsay 1999 p.321, Jensen et al. 2006).

En particular, el Supuesto A3 es requerido para la estimación puntual y por intervalos de confianza del PCI, mientras que se requiere el cumplimiento de los Supuestos A1 y A2 para la construcción de intervalos de confianza del PCI. A modo de ejemplo, el Supuesto A2 se cumple si las muestras son seleccionadas de una población infinita.

Asumiendo el Supuesto A1 y el Supuesto A2, el Supuesto A3 implica que ambos parámetros de la distribución normal ( $\mu$  y  $\sigma$ ) están bajo control, es decir, el estadístico  $S$  se sitúa dentro de los límites de control del gráfico de control usado para controlar la variabilidad del proceso:

$$\begin{aligned} LIC_\sigma &= \left( c_4[n] - c\sqrt{1 - c_4^2[n]} \right) \sigma \\ LSC_\sigma &= \left( c_4[n] + c\sqrt{1 - c_4^2[n]} \right) \sigma \end{aligned} \quad (2.4)$$

donde  $LIC_\sigma$  denota el límite inferior de control,  $LSC_\sigma$  es el límite superior de control y  $c_4[n]$  es una función que depende del tamaño de la muestra (véase Chen 1997 o el Capítulo 3 de este trabajo). Algunos valores de  $c_4[n]$  para varios

tamaños muestrales han sido tabulados en el Apéndice VI de Montgomery (2009).  $c$  es una constante usada en los gráficos de control para definir la distancia de los límites de control respecto a la línea central del gráfico de control, que es fijada en

$$LC_\sigma = c_4[n]\sigma$$

en este caso. Destacamos que es bastante común considerar  $c = 3$  (ver Montgomery 2009 p.181, Evans y Lindsay 1999 p.321, Jensen et al. 2006). De forma similar, el parámetro  $\mu$  está bajo control si el estadístico  $\bar{x}$  se sitúa dentro de los límites de control del gráfico de control para la media del proceso:

$$\begin{aligned} LIC_\mu &= \mu - c \frac{\sigma}{\sqrt{n}} \\ LSC_\mu &= \mu + c \frac{\sigma}{\sqrt{n}} \end{aligned} \tag{2.5}$$

donde se suele considerar el valor  $c = 3$ . En este caso, la línea central de este gráfico viene dada por

$$LC_\mu = \mu$$

Destacamos que en la práctica existe cierta relación entre el Supuesto A1 y el Supuesto A3. En particular, el incumplimiento del Supuesto A3 puede afectar al cumplimiento del Supuesto A1. De forma genérica, un proceso productivo debe estar calibrado de tal manera que produzca artículos que presenten un valor objetivo en su característica de calidad. Si dicho proceso se encuentra en una situación definida por los gráficos de control como bajo control (Supuesto A3), su variabilidad solo será atribuible a causas fortuitas. Por tanto, esa variabilidad se traducirá en que la producción de artículos presenta un valor en su característica de calidad cercano, tanto por defecto como por exceso, al valor objetivo. En este sentido, se puede justificar el hecho de asumir, en numerosas situaciones, que la variable  $X$  se distribuya según una distribución Normal (Supuesto A1), con media, presumiblemente, cercana al valor objetivo y una desviación típica que representa la variabilidad producida por las causas fortuitas. En el lado opuesto, si un proceso calibrado tal como se ha argumentado está en un estado definido como fuera de control (incumplimiento del Supuesto A3), la variabilidad presente en éste será atribuible a causas asignables. Este hecho puede implicar que el valor medio de la característica de

calidad de los artículos difiera en gran medida del valor objetivo de la misma. Sin embargo, las observaciones tomadas de los artículos en los que la causa asignable no está presente si pueden proceder de una distribución Normal, y sería la presencia de los artículos producidos bajo causas asignables los que pueden alterar la distribución, y por tanto el incumplimiento del Supuesto A3. En otras palabras, en numerosas ocasiones el incumplimiento del Supuesto A1 sobre normalidad puede deberse a causas asignables tal como el desgaste de alguna pieza del proceso de producción, cambio de turno de los trabajadores, cambio significativo de las condiciones ambiental, etc. En esta situación, el incumplimiento del Supuesto A1 se debe al incumplimiento del Supuesto A3, es decir, al realizar los gráficos de control se detectaría que el proceso está fuera de control (se incumple el Supuesto A3), y lo que procede realizar es la detección de las causas asignables, eliminar de la muestra los productos producidos bajo estas circunstancias, y realizar acciones correctivas en el proceso de producción para tenerlo bajo control. Con el seguimiento de estas pautas se podría garantizar el cumplimiento de los Supuestos A3 y A1. A continuación se exponen dos casos prácticos que permiten explicar de forma más detallada las situaciones descritas.

Supongamos una empresa que opera con un proceso en el cual interviene una determinada máquina compuesta por una serie de engranajes. Un engranaje en particular contiene un diente de unas dimensiones no adecuadas, consecuencia de un desgaste excesivo. Cada cierto número de artículos producidos por el proceso, el diente defectuoso de dicha máquina interviene en él. Supongamos que este hecho influye directamente en la característica de calidad del producto, dando como resultado un valor superior al habitual. Si se elabora un gráfico de control del proceso, se podría observar que los valores del estadístico correspondientes a muestras que incluyen artículos producidos con la intervención del diente defectuoso serán, en general, más elevados que los valores del estadístico del resto de muestras. Dicho punto se podría situar por encima del *LSC* o, simplemente, representar un patrón de comportamiento en el gráfico de control, y ambos casos pueden ser un indicio de que el proceso no está bajo control. En consecuencia, podría afirmarse que el proceso se encuentra en un estado de fuera de control. Es decir, una parte de la variabilidad del proceso es producida por una causa asignable: el diente de dimensiones no adecuadas. Esta situación vulnera el Supuesto A3 (el proceso está bajo control) y, posiblemente, también el Supuesto A1 (la característica de calidad procede de una distribución Normal con parámetros  $\mu$  y  $\sigma$ ).

Analicemos también a continuación el siguiente caso. Una empresa trabaja con un proceso en el cual interviene una máquina que, tras producir un determinado número de artículos, se sobrecalienta. Mientras dicha máquina está

sobrecalentada, el proceso produce artículos con una alteración determinada en su característica de calidad, tal como una longitud ligeramente mayor. Si se elabora el gráfico de control correspondiente a este proceso podría identificarse en la serie cronológica del estadístico muestral un patrón, una racha, una serie más o menos larga de puntos situados a una determinada distancia de la línea central del gráfico de control, o incluso, puntos situados fuera del intervalo definido por los límites de control. Todos estos escenarios están contemplados entre las reglas comunes (véase la Sección 1.1.1) que establecen que el gráfico está fuera de control. Por tanto, al igual que en la situación anterior, podrían observarse indicios que indiquen que el proceso se encuentra en una situación de fuera de control. En este caso, la causa asignable es la máquina defectuosa. Del mismo modo, se incumple el Supuesto A3 y, probablemente como en el caso anterior, el Supuesto A1.

Para ilustrar estos casos prácticos, a continuación se presentan algunos ejemplos, en los cuales se elaboran los correspondientes gráficos de control para la media y la desviación típica, y asumiendo, por simplicidad, que los parámetros son conocidos para los distintos procesos. En cada uno de los procesos de los ejemplos presentados, la producción se almacena en lotes, de los cuales, se extrae una muestra de artículos. En cada artículo se mide la característica de calidad y se obtiene el estadístico asociado a cada muestra. Posteriormente, se elabora el gráfico de control. Se contemplan tres escenarios entre los ejemplos presentados. El primer escenario corresponde a un proceso productivo que opera de forma correcta. En el segundo escenario, el proceso productivo opera con una cierta alteración en la característica de calidad, y que repercute sólo en una proporción  $p^*$  de los artículos producidos. En este segundo escenario, asumiremos que cada lote contiene dicha proporción  $p^*$  de artículos. Por último, el tercer escenario corresponde a un proceso productivo que, a partir de un cierto volumen de producción, cambia su rendimiento. Puesto que el almacenamiento en lotes es cronológico, suponemos que sólo un número determinado de lotes contiene artículos producidos por el proceso cuando presenta la incidencia. Además, supondremos que la proporción de lotes con artículos con este tipo de incidencias también viene dada por  $p^*$ . Por ejemplo, si suponemos  $p^* = 0,2$  y lotes de tamaño 100, en el segundo escenario cada lote almacena 20 artículos con incidencia. Bajo el tercer escenario y de manera cíclica, el 20 % de los lotes contendrán artículos con incidencias, es decir, la producción se irá almacenando en cuatro lotes consecutivos con artículos sin incidencia, mientras que el siguiente lote contendrá artículos con incidencia. Destacamos que los casos prácticos expuestos anteriormente pueden identificarse con el segundo y tercer escenario, respectivamente.

Por tanto, se presentan dos tipos distintos de producción. Por un lado, la

producción correspondiente a un proceso cuando opera sin ninguna incidencia y, por otro, la producción resultado de un proceso cuando es afectado por alguna incidencia. Asumimos que la variable que representa la característica de calidad de los artículos producidos por un proceso sin incidencias tiene distribución una normal con media  $\mu$  y desviación típica  $\sigma$ . Por su parte, la característica de calidad correspondiente a la producción con incidencias sigue una distribución normal con media  $\mu + \epsilon_\mu$  y desviación típica  $\sigma + \epsilon_\sigma$ . En los ejemplos que se exponen a continuación se han elaborado gráficos de control en los que se representan 50 muestras de tamaño 5 y los límites de control son definidos usando el método USA, es decir,  $c = 3$ . Para simular los 3 escenarios definidos con anterioridad se han asumido lotes de 100 artículos de capacidad y una la proporción de incidencias del 20% ( $p^* = 0,2$ ) para el segundo y tercer escenario. Respecto a la producción, se ha asumido que los valores de los parámetros son  $\mu = 10$  y  $\sigma = 1$ . Además, se asumen los valores  $\epsilon_\mu = \{0, 1, 2, 3\}$  y  $\epsilon_\sigma = \{0, 0,5, 1, 2\}$ , y todas sus posibles combinaciones, con el fin de representar las alteraciones en la producción comentadas.

Destacamos que la idea principal de este estudio, además de ilustrar y exponer la relación entre los Supuestos A1 y A3, es evaluar el impacto que tiene en el proceso incidencias con diferente repercusión, tanto en intensidad como en forma. Los principales resultados de la simulación descrita se muestran en las Figuras 2.1-2.9. Cada Figura contiene los tres escenarios planteados. Destacamos que el primer escenario se representa en todas las figuras a modo de referencia, ya que éste se caracteriza por ser un proceso bajo control que no sufre variación.

Las Figuras 2.1 y 2.2 contienen, respectivamente, los gráficos de control para la media y la desviación típica cuando la incidencia en el proceso tiene influencia únicamente en la media de la característica de calidad, es decir,  $\epsilon_\mu = 1$  y  $\epsilon_\sigma = 0$ . En este caso, se observa en el gráfico de control para la media (Figura 2.1) indicios suficientes para pensar que, tanto en el segundo como en el tercer escenario, el proceso está fuera de control. En el segundo escenario, la media de la muestra número 17 está fuera del rango delimitado por los límites de control. Algo similar se observa en las muestras número 5 y 45 en el tercer escenario. Por su parte, los gráficos de la desviación típica de la Figura 2.2 no muestra indicios para pensar que el parámetro está fuera de control en ninguno de los escenarios. Los gráficos de control para la media y la desviación típica cuando la incidencia en el proceso influye únicamente a la media de la característica de calidad de forma severa ( $\epsilon_\mu = 2$  y  $\epsilon_\sigma = 0$ ), se muestra respectivamente en las Figuras 2.3 y 2.4. Se observa que, en el tercer escenario, prácticamente la totalidad de las medias muestrales correspondientes a la producción con incidencia en el proceso se sitúan por encima del *LSC*,

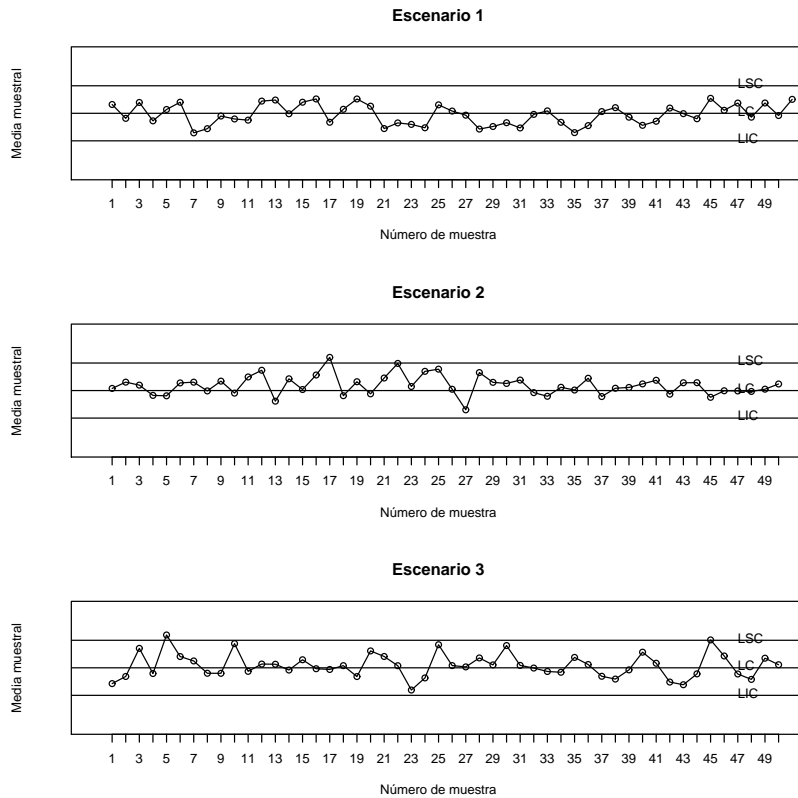


Figura 2.1: Gráfico de control para la media. Se seleccionan muestras de una distribución Normal  $\mu = 10$  y  $\sigma = 1$ . La producción con incidencia representa una proporción  $p^* = 0,2$ , y su característica de calidad está afectada por los valores  $\epsilon_\mu = 1$  y  $\epsilon_\sigma = 0$ .

mientras que la desviación típica parece encontrarse bajo control. Respecto al proceso del segundo escenario, puede afirmarse que se encuentra fuera de control, ya que, la desviación típica de la muestra número 16 se sitúa fuera del intervalo definido por los límites de control. Además, se observa en las muestras número 5 y 6 otra situación para identificar el proceso como fuera de control, “dos de tres puntos consecutivos fuera de los límites de advertencia de dos sigma pero aún dentro de los límites de control”.

Las Figuras 2.5 y 2.8 muestran los gráficos de control cuando la incidencia en el proceso solo afecta a la desviación típica de la característica de calidad del producto. Cuando la incidencia influye de manera relativamente moderada, es decir,  $\epsilon_\mu = 0$  y  $\epsilon_\sigma = 1$  (Figuras 2.5 y 2.6), se observan indicios de que el



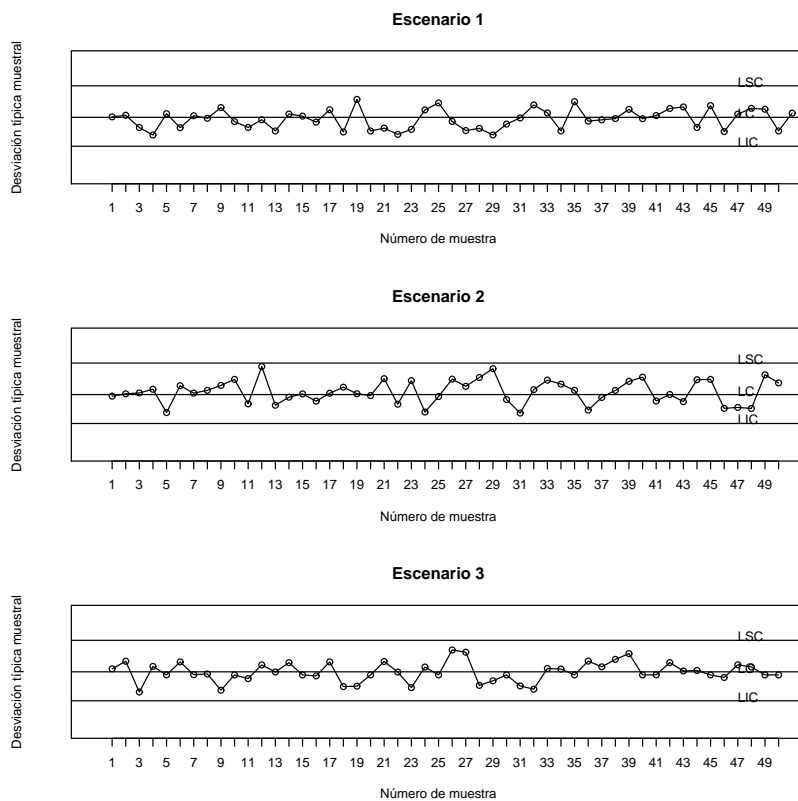


Figura 2.2: Gráfico de control para la desviación típica. Se seleccionan muestras de una distribución Normal  $\mu = 10$  y  $\sigma = 1$ . La producción con incidencia representa una proporción  $p^* = 0,2$ , y su característica de calidad está afectada por los valores  $\epsilon_\mu = 1$  y  $\epsilon_\sigma = 0$ .

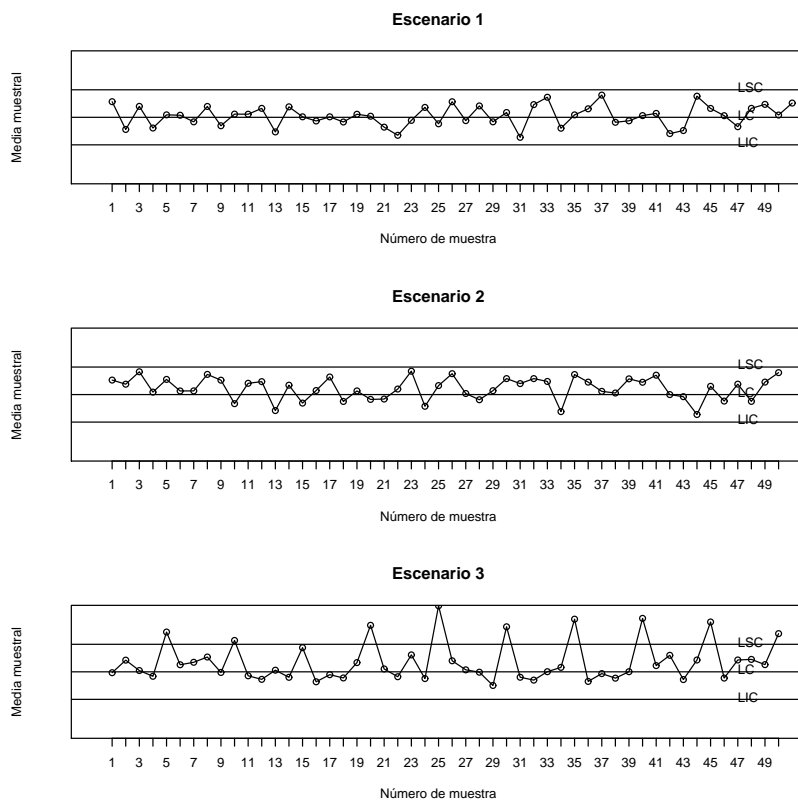


Figura 2.3: Gráfico de control para la media. Se seleccionan muestras de una distribución Normal  $\mu = 10$  y  $\sigma = 1$ . La producción con incidencia representa una proporción  $p^* = 0,2$ , y su característica de calidad está afectada por los valores  $\epsilon_\mu = 2$  y  $\epsilon_\sigma = 0$ .

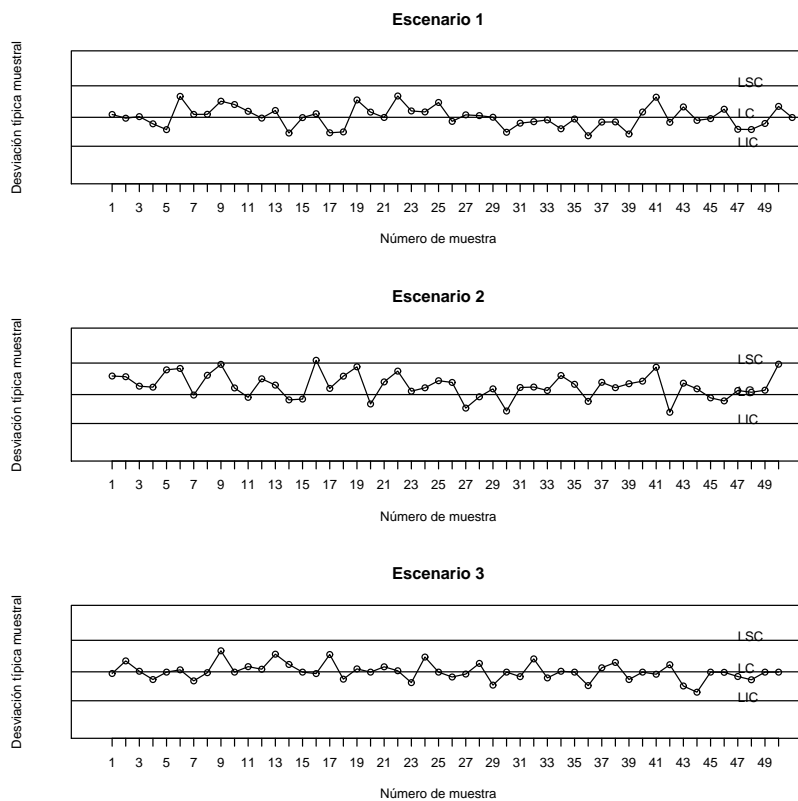


Figura 2.4: Gráfico de control para la desviación típica. Se seleccionan muestras de una distribución Normal  $\mu = 10$  y  $\sigma = 1$ . La producción con incidencia representa una proporción  $p^* = 0,2$ , y su característica de calidad está afectada por los valores  $\epsilon_\mu = 2$  y  $\epsilon_\sigma = 0$ .

proceso correspondiente al segundo y tercer escenario se encuentran fuera de control. En el gráfico de control de la desviación típica correspondiente al proceso del segundo escenario, se observan que el estadístico de las muestras número 8, 14 y 17 están fuera de los límites de control. También se detecta una racha descendente desde la muestra número 20 hasta la 25, situación que está contemplada como indicio de que el proceso está en una situación de fuera de control. Respecto al proceso del tercer escenario, el gráfico de control de la media presenta un punto correspondiente a la muestra número 10 fuera de los límites de control, por lo que hay indicios de que el proceso esté fuera de control. En el caso de que la incidencia afectara de forma severa la desviación típica de la característica de calidad, por ejemplo,  $\epsilon_\mu = 0$  y  $\epsilon_\sigma = 2$  (Figuras 2.7 y 2.8) esta situación se hace más evidente. Por ejemplo, en el segundo escenario, el gráfico de control de la desviación típica contiene el estadístico correspondientes a varias muestras muy alejadas del intervalo definido por los límites de control. El gráfico de control de la media también presenta esta situación y, además, se aprecia una racha de siete muestras consecutivas donde el valor del estadístico disminuye. Los gráficos de control correspondientes al proceso del tercer escenario también presentan estas características. Por ejemplo, la muestra número 5 presenta un valor para la media muestral por encima del *LSC* y, desde la muestra 21 hasta la 26, una racha descendente en el valor de la desviación típica muestral.

Por último, los gráficos de control para la media y la desviación típica asumiendo valores  $\epsilon_\mu = 1$  y  $\epsilon_\sigma = 1$  se muestran en las Figuras 2.9 y 2.10, respectivamente. En este caso la incidencia en el proceso afecta tanto a la media como a la desviación típica de la característica de calidad del producto. Los gráficos de control contienen indicios de que los procesos del segundo y tercer escenario se encuentran fuera de control. Entre éstos, destaca la existencia de puntos fuera de los límites de control del gráfico de control para la media en ambos procesos, o puntos fuera de los límites de control en el gráfico de la desviación típica correspondiente al proceso del segundo escenario.

Como se indicó con anterioridad, los gráficos de control monitorizan el proceso de producción para detectar causas asignables, lo que posibilita aplicar acciones correctivas sobre ellas. En los casos comentados, estas acciones podrían ser: sustituir el engranaje con el diente defectuoso y corregir el problema de refrigerado de la máquina con problemas de sobrecalentamiento. Una vez eliminadas las causas asignables, podría resultar posible pasar el proceso de producción a un estado de bajo control (cumpliéndose el Supuesto A3), e incluso que la característica de calidad  $X$  se distribuya según una distribución Normal (cumpliéndose el Supuesto A1).

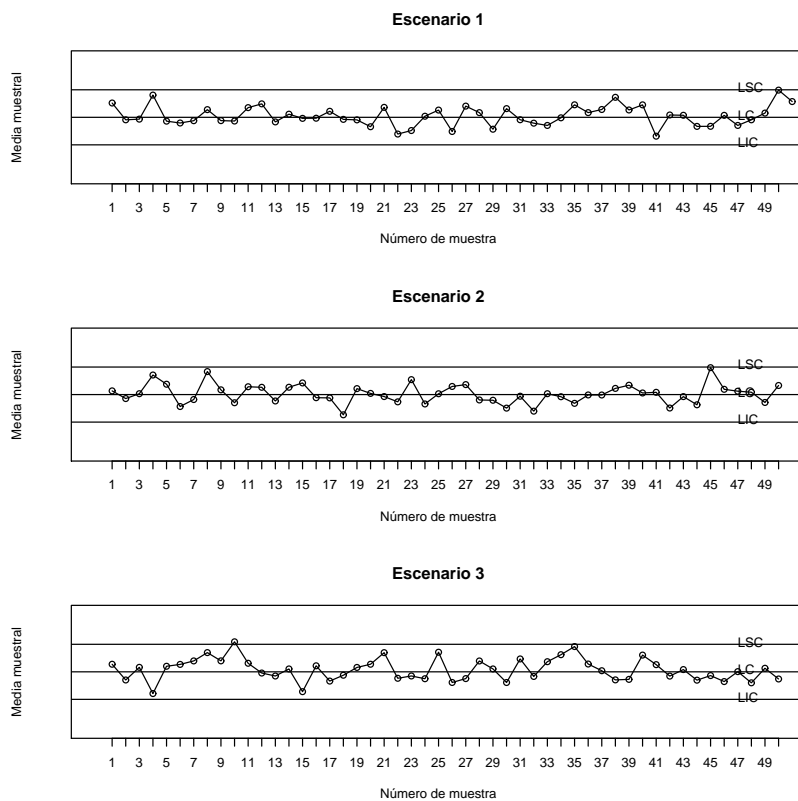


Figura 2.5: Gráfico de control para la media. Se seleccionan muestras de una distribución Normal  $\mu = 10$  y  $\sigma = 1$ . La producción con incidencia representa una proporción  $p^* = 0,2$ , y su característica de calidad está afectada por los valores  $\epsilon_\mu = 0$  y  $\epsilon_\sigma = 1$ .

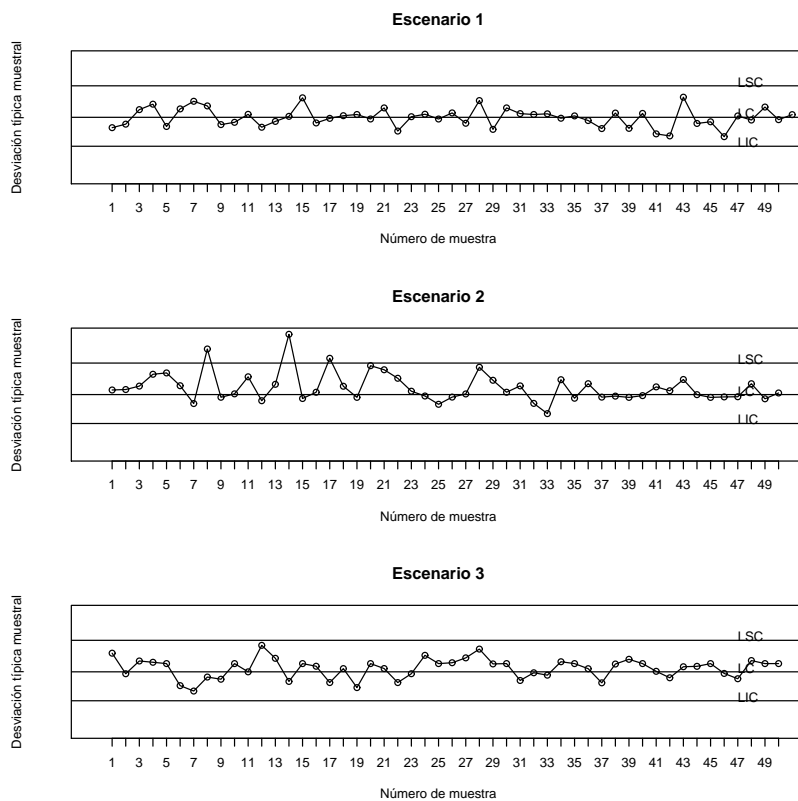


Figura 2.6: Gráfico de control para la desviación típica. Se seleccionan muestras de una distribución Normal  $\mu = 10$  y  $\sigma = 1$ . La producción con incidencia representa una proporción  $p^* = 0,2$ , y su característica de calidad está afectada por los valores  $\epsilon_\mu = 0$  y  $\epsilon_\sigma = 1$ .

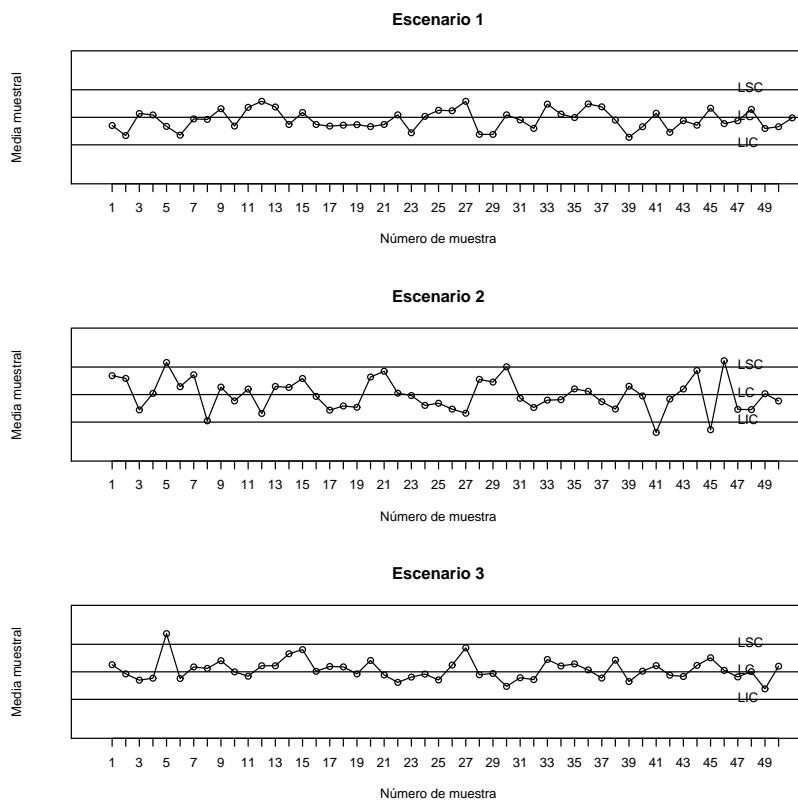


Figura 2.7: Gráfico de control para la media. Se seleccionan muestras de una distribución Normal  $\mu = 10$  y  $\sigma = 1$ . La producción con incidencia representa una proporción  $p^* = 0,2$ , y su característica de calidad está afectada por los valores  $\epsilon_\mu = 0$  y  $\epsilon_\sigma = 2$ .

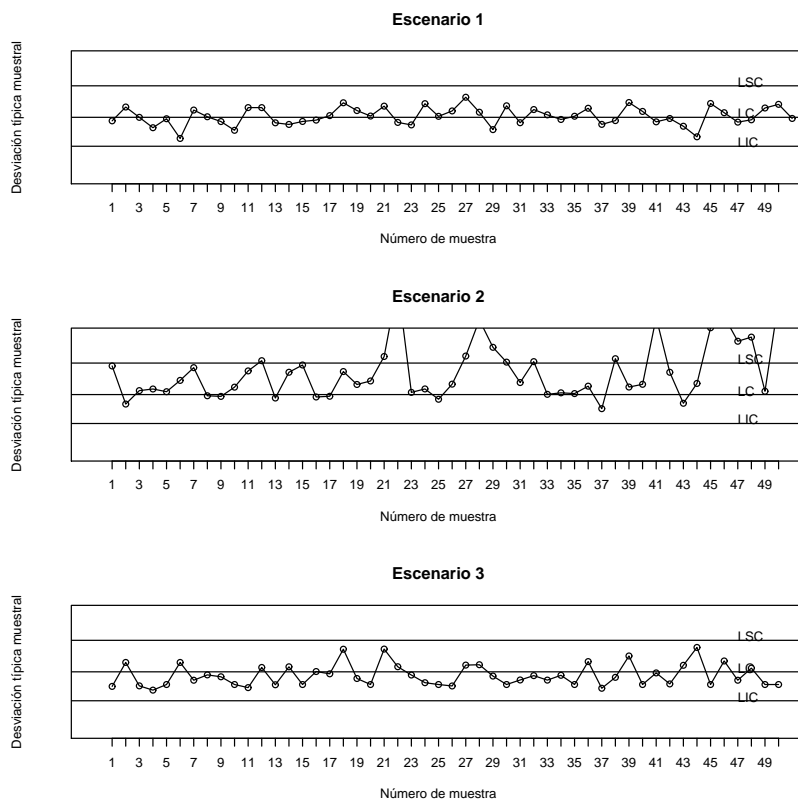


Figura 2.8: Gráfico de control para la desviación típica. Se seleccionan muestras de una distribución Normal  $\mu = 10$  y  $\sigma = 1$ . La producción con incidencia representa una proporción  $p^* = 0,2$ , y su característica de calidad está afectada por los valores  $\epsilon_\mu = 0$  y  $\epsilon_\sigma = 2$ .



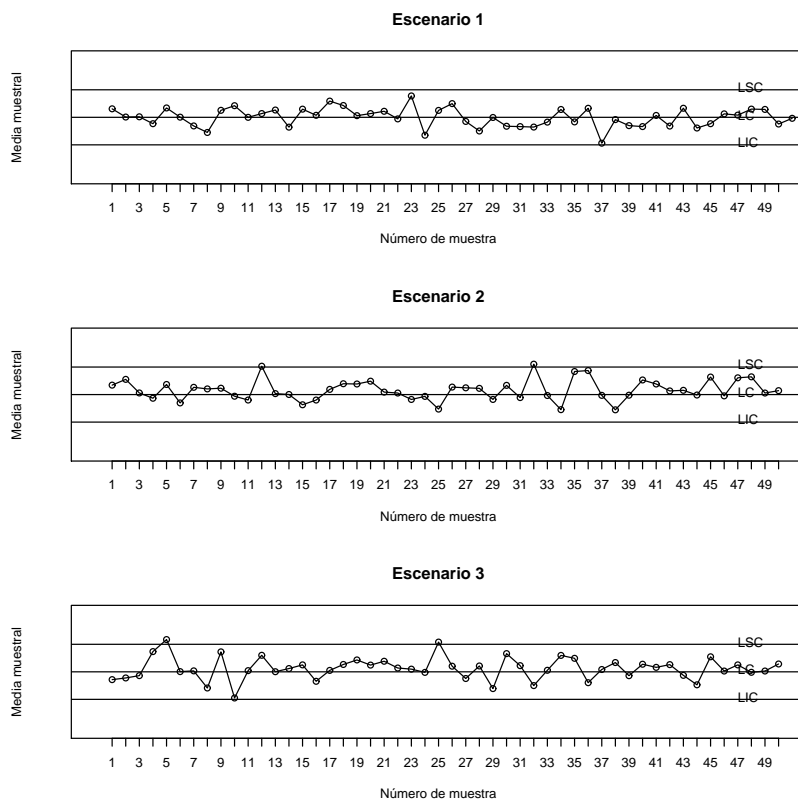


Figura 2.9: Gráfico de control para la media. Se seleccionan muestras de una distribución Normal  $\mu = 10$  y  $\sigma = 1$ . La producción con incidencia representa una proporción  $p^* = 0,2$ , y su característica de calidad está afectada por los valores  $\epsilon_\mu = 1$  y  $\epsilon_\sigma = 1$ .

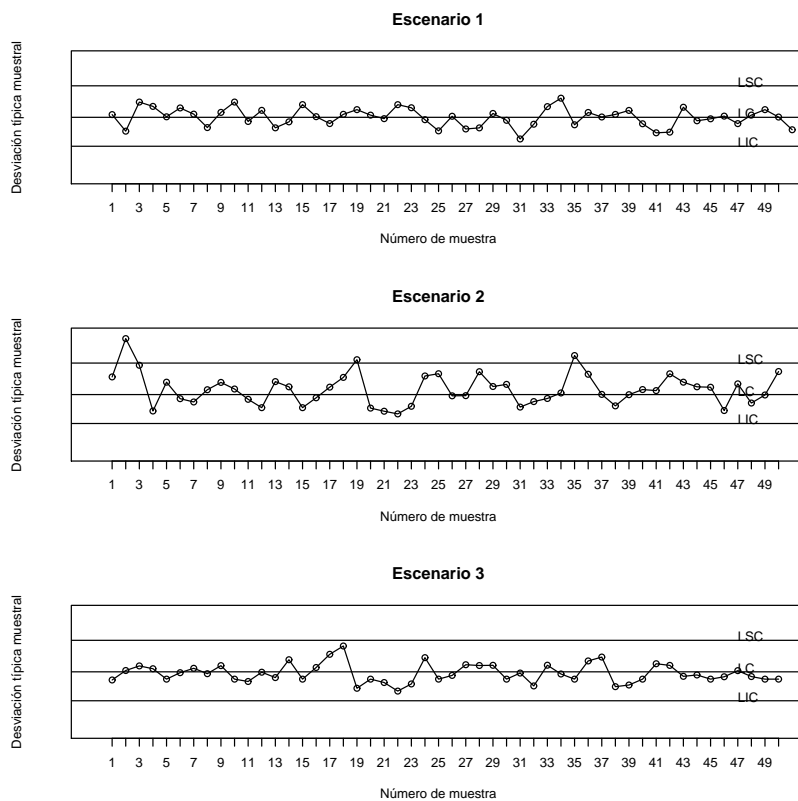


Figura 2.10: Gráfico de control para la desviación típica. Se seleccionan muestras de una distribución Normal  $\mu = 10$  y  $\sigma = 1$ . La producción con incidencia representa una proporción  $p^* = 0,2$ , y su característica de calidad está afectada por los valores  $\epsilon_\mu = 1$  y  $\epsilon_\sigma = 1$ .

## 2.2.2. Estimación puntual y por intervalos de confianza

### Estimación puntual

La desviación típica del proceso  $\sigma$  es habitualmente desconocida, y por tanto, el PCI definido por la ecuación (2.1) también será desconocido. Un estimador para  $C_p$  se obtiene sustituyendo  $\sigma$  por su estimador en la expresión (2.1). Se consideran los estimadores para  $\sigma$  basados en la desviación típica muestral, ya que el intervalo de confianza para el PCI más común está basado en este estimador. El estimador usual de  $\sigma$  es definido por:

$$\hat{\sigma} = S$$

Como es bien sabido, la cuasivarianza muestral,  $\hat{\sigma}^2 = S^2$ , es un estimador insesgado de la varianza poblacional  $\sigma^2$ . Sin embargo, la raíz cuadrada no es una función lineal, y en consecuencia  $\hat{\sigma}$  no es un estimador insesgado de  $\sigma$ . Por esta razón, y asumiendo el Supuesto A2, es más común usar el estimador

$$\hat{\sigma}_{c_4} = \frac{S}{c_4[n]}$$

especialmente para tamaños muestrales pequeños. La función  $c_4[n]$  depende del tamaño de muestra  $n$ , y su definición puede verse también en Chen (1997). Los valores de  $c_4[n]$  están también tabulados para diferentes tamaños muestrales en el Appendix VI de Montgomery (2009). Destacar que  $c_4[n]$  se usa porque proporciona un estimador insesgado de la desviación típica del proceso  $\sigma$  (Duncan 1986, Luko 1996, Wheeler 1995, y Woodall y Montgomery 2000). Según se indica en el Capítulo 3 de este trabajo, no puede obtenerse el valor de la función  $c_4[n]$  para tamaños muestrales elevados y siguiendo la definición de esta función. Por este motivo principalmente, se proporcionan nuevas expresiones para  $c_4[n]$ , y las cuales si permiten su cálculo para cualquier valor de  $n$ . Destacamos que en numerosos textos sobre *SQC* se habla de la "constante"  $c_4$ . En efecto, se utiliza en el caso de fijar el tamaño muestral.

Siguiendo la ecuación (2.1), los estimadores del índice de capacidad  $C_p$  basados en los estimadores  $\hat{\sigma}$  y  $\hat{\sigma}_{c_4}$  se definen, respectivamente, como:

$$\hat{C}_p = \frac{LSE - LIE}{6\hat{\sigma}}, \quad (2.6)$$

$$\widehat{C}_{p.c4} = \frac{LSE - LIE}{6\widehat{\sigma}_{c4}}, \quad (2.7)$$

La precisión de los estimadores de  $\sigma$  es un tema ampliamente discutido en la literatura. El sesgo y la eficiencia de diferentes estimadores de  $\sigma$  son tratados por Chakraborti (2008), Chen (1997), Duncan (1986), Jones et al. (2001), Luko (1996), Montgomery (2009, pp. 229 y 253), Ott (1975), Vanderman (1999), Wheeler (1995), Woodall and Montgomery (2000), etc. A continuación (Resultado 2.1) se analiza desde un punto de vista teórico el sesgo del estimador  $\widehat{C}_{p.c4}$ , el cual está basado en  $\widehat{\sigma}_{c4}$ , un estimador insesgado de  $\sigma$ .

**Resultado 2.1** El sesgo relativo del estimador  $\widehat{C}_{p.c4}$  del índice de capacidad, definido por

$$BR(\widehat{C}_p) = \frac{B(\widehat{C}_{p.c4})}{\{V(\widehat{C}_{p.c4})\}^{1/2}},$$

está acotado superiormente por el error estándar relativo del estimador  $\widehat{\sigma}_{c4}$ , el cual viene definido por

$$RSE(\widehat{\sigma}_{c4}) = \frac{\{V(\widehat{\sigma}_{c4})\}^{1/2}}{\sigma},$$

es decir

$$BR(\widehat{C}_{p.c4})^2 \leq RSE(\widehat{\sigma}_{c4})^2, \quad (2.8)$$

y donde  $B(\widehat{C}_{p.c4})$  y  $V(\widehat{C}_{p.c4})$  denotan, respectivamente, el sesgo y la varianza de  $\widehat{C}_{p.c4}$ .  $\square$

### Demostración

El estimador  $\widehat{C}_{p.c4}$  puede expresarse como:

$$\widehat{C}_{p.c4} = \frac{USL - LSL}{6\widehat{\sigma}} = \frac{L^*}{\widehat{\sigma}_{c4}}, \quad (2.9)$$

donde

$$L^* = \frac{USL - LSL}{6}.$$

De la expresión (2.9) se obtiene que  $\widehat{C}_{p.c4}\widehat{\sigma}_{c4} = L^*$ , y puesto que  $\widehat{\sigma}_{c4}$  es insesgado, la covarianza entre  $\widehat{C}_{p.c4}$  y  $\widehat{\sigma}_{c4}$  viene dada por

$$Cov[\widehat{C}_{p.c4}, \widehat{\sigma}_{c4}] = E[\widehat{C}_{p.c4}\widehat{\sigma}_{c4}] - E[\widehat{C}_{p.c4}]E[\widehat{\sigma}_{c4}] = E[L^*] - E[\widehat{C}_{p.c4}]E[\widehat{\sigma}_{c4}] =$$

$$= L^* - E[\widehat{C}_{p,4}]\sigma = C_p\sigma - E[\widehat{C}_{p,c4}]\sigma = -\sigma(E[\widehat{C}_{p,c4}] - C_p) = -\sigma B(\widehat{C}_{p,c4})$$

En otras palabras, el sesgo de  $\widehat{C}_p$  puede ser expresado como

$$B(\widehat{C}_{p,c4}) = E[\widehat{C}_{p,c4}] - C_p = -\frac{Cov[\widehat{C}_{p,c4}, \widehat{\sigma}_{c4}]}{\sigma}$$

Entonces, se deduce

$$B(\widehat{C}_{p,c4})^2 = \frac{Cov[\widehat{C}_{p,c4}, \widehat{\sigma}_{c4}]^2}{\sigma^2} = \frac{\rho(\widehat{C}_{p,c4}, \widehat{\sigma}_{c4})^2 \cdot V(\widehat{C}_{p,c4}) \cdot V(\widehat{\sigma}_{c4})}{\sigma^2},$$

donde  $\rho(\widehat{C}_p, \widehat{\sigma})$  es el coeficiente de correlación entre  $\widehat{C}_p$  y  $\widehat{\sigma}$ , cuyo cuadrado está acotado superiormente por la unidad, y por tanto se puede llegar a la siguiente expresión

$$B(\widehat{C}_{p,c4})^2 \leq \frac{V(\widehat{C}_{p,c4}) \cdot V(\widehat{\sigma}_{c4})}{\sigma^2},$$

que es similar a

$$\frac{B(\widehat{C}_{p,c4})^2}{V(\widehat{C}_{p,c4})} = BR(\widehat{C}_{p,c4})^2 \leq RSE(\widehat{\sigma}_{c4})^2.$$

□

Se puede observar que  $\widehat{C}_{p,c4}$  se ha expresado como el cociente entre la constante  $L^*$  y el estimador  $\widehat{\sigma}_{c4}$ . Aunque el estimador  $\widehat{\sigma}_{c4}$  es insesgado, destacamos que el estimador  $\widehat{C}_{p,c4}$  no tiene esta propiedad, es decir,  $E[\widehat{C}_{p,c4}] \neq C_p$  (véase también Särndal et al. 1992 p.163). El Resultado 2.1 indica que si  $RSE(\widehat{\sigma}_{c4})$  se aproxima a cero cuando se aumenta el tamaño de la muestra, el sesgo estándar de  $\widehat{C}_{p,c4}$  también se aproximará a cero. Debe tenerse en cuenta que es bastante común tener errores estándar relativos cercanos a cero cuando el tamaño de la muestra es elevado, y por tanto, el sesgo estándar de  $\widehat{C}_{p,c4}$  será pequeño en esta situación (véase también Särndal et al. 1992 p.176).

Con el objetivo de ilustrar y analizar en distintos escenarios el Resultado 2.1, se ha realizado un estudio de simulación basado en 1000 repeticiones, y en el cual se determina a través de éstas el valor medio del cuadrado del error estándar relativo de los estimadores  $\widehat{C}_{p.c4}$  y  $\widehat{C}_p$ . En este estudio, se han tenido en cuenta datos muestrales de las distribuciones Normal, Gamma y Uniforme. Además, se han considerado muestras que contienen el 5 % y el 20 % de valores atípicos, con el objetivo de estudiar el impacto de la presencia de dichos valores extremos. En la Figura 2.11 se muestran los resultados de la simulación para diferentes tamaños muestrales. Como se indicó con anterioridad, se observa que la cota superior para el sesgo relativo del estimador  $\widehat{C}_{p.c4}$  se aproxima a cero a medida que el tamaño de muestra aumenta. En el caso del estimador  $\widehat{C}_p$  se observa que, en general, los valores medios del correspondiente cuadrado del error estándar relativo es superior a los obtenidos para el estimador  $\widehat{C}_{p.c4}$  y para tamaños muestrales menores que 100. Por otra parte, se puede observar que la inclusión de datos atípicos no tiene un impacto relevante en la cota superior del estimador  $\widehat{C}_{p.c4}$ , por lo que el sesgo tampoco se verá afectado por la presencia de datos anómalos.

Se puede observar que el estimador  $\widehat{C}_p$  definido por la expresión (2.9) es una función no lineal de las observaciones. Cabe destacar que las varianzas de estadísticos complejos, tales como  $\widehat{C}_p$ , no pueden ser expresadas mediante fórmulas teóricas sencillas (veáse también Rueda y Muñoz 2011). Además, Wolter (2007, p. 119) indica que para la estimación de la varianza de estadísticos no lineales, sólo se pueden proporcionar, por lo general, resultados aproximados, es decir, hay una escasez de resultados teóricos exactos para el escenario de muestras seleccionadas de una población finita. En el caso de los estadísticos complejos o no lineales es bastante común el uso de técnicas tradicionales tales como jackknife (Särndal et al. 1992 p.437, Wolter 2007 p.151) o bootstrap (Särndal et al. 1992 p.442, Wolter 2007 p.194) para estimar la varianza de los estimadores correspondientes. Por ejemplo, como se comenta por Wolter (2007, p.119), numerosos resultados empíricos sugieren que los métodos de remuestreo ofrecen estimaciones deseables de la verdadera varianza de un estimador definido por un ratio. En consecuencia, se aconseja utilizar técnicas tradicionales de remuestreo para estimar la varianza de  $\widehat{C}_p$ , puesto que ésta es una solución simple que puede proporcionar resultados satisfactorios. Además, muchos programas estadísticos incluyen paquetes y herramientas que implementan métodos de aproximación de la varianza, y de ahí que el uso de ellos en la práctica es bastante simple.

Para procesos no centrados, los estimadores puntuales del correspondiente índice de capacidad  $C_{pk}(PCI_k)$  siguen las siguientes expresiones:

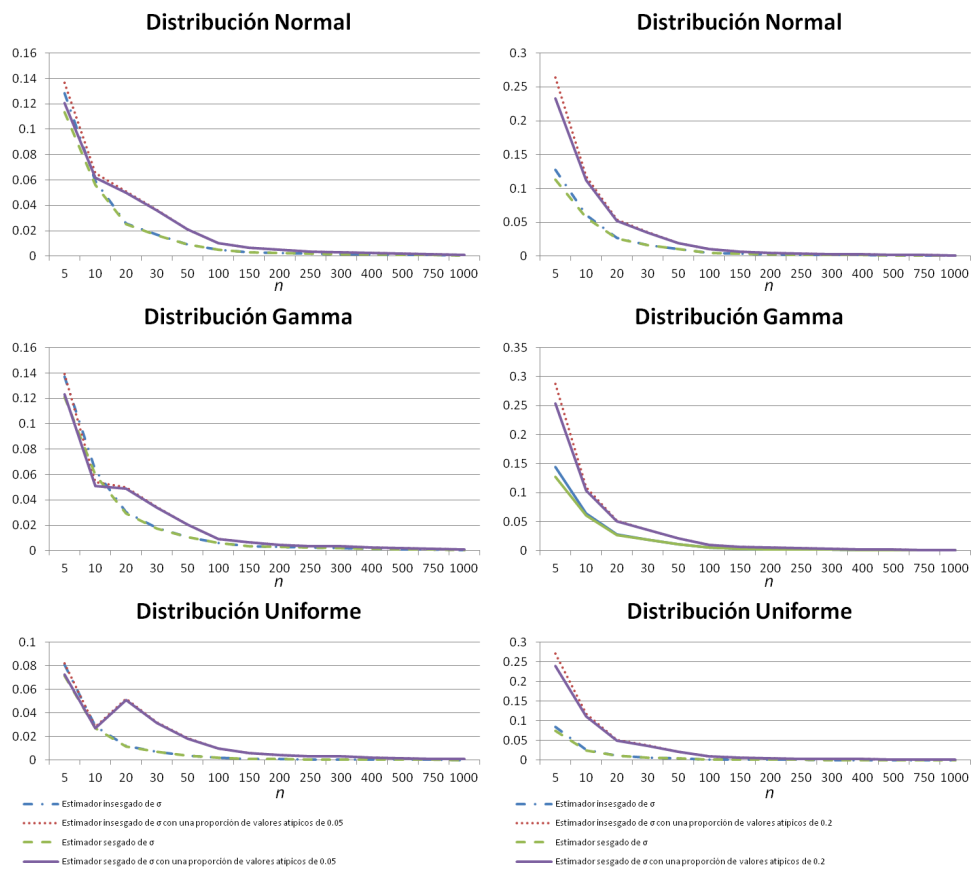


Figura 2.11: Valores medios del cuadrado del error estándar relativo de los estimadores  $\hat{C}_{p.c4}$  y  $\hat{C}_p$ , y basados en datos procedentes de las distribuciones Normal, Gamma y Uniforme. También se incluyen escenarios con proporciones del 5 % y del 20 % de valores atípicos en la muestra.

$$\widehat{C}_{pk} = (1 - \widehat{k})\widehat{C}_p$$

y

$$\widehat{C}_{pk.c4} = (1 - \widehat{k})\widehat{C}_{p.c4}$$

donde

$$\widehat{k} = \frac{|\mu_0 - \widehat{\mu}|}{(LSE - LIE)/2},$$

y  $\widehat{\mu} = \bar{x}$  es la media muestral.

De las expresiones (2.6) y (2.7) se observa que la precisión de la estimación del PCI está directamente relacionada con la precisión de la estimación de la desviación típica del proceso. En este capítulo se investiga el comportamiento empírico de los estimadores  $\widehat{\sigma}$  y  $\widehat{\sigma}_{c4}$ , y tales resultados son comparados con el rendimiento empírico de los estimadores de PCI ( $\widehat{C}_p$  y  $\widehat{C}_{p.c4}$ ) y  $PCI_k$  ( $\widehat{C}_{pk}$  y  $\widehat{C}_{pk.c4}$ ).

### Estimación mediante intervalos de confianza

La construcción de intervalos de confianza para  $PCI$  y  $PCI_k$  ha sido extensamente investigada (Kane 1986, Heavlin 1988, Chou et al. 1990, Chou y Owen 1989, Lie et al. 1990). Sin embargo, tales referencias asumen los Supuestos A1, A2 y A3. En particular, es bastante conocido que bajo estas condiciones se cumple

$$\frac{(n-1)\widehat{\sigma}^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2 \equiv \gamma\left(\frac{n-1}{2}, 2\right),$$

Si se considera el estimador  $\widehat{C}_p$  definido por (2.6), y se usa la distribución de la expresión anterior, se deduce fácilmente que un intervalo de confianza, al  $(1 - \alpha)\%$  para el nivel de confianza, para  $PCI$  se define como (Kotz y Johnson 1993):

$$\left[ \frac{\chi_{n-1, \alpha/2}}{\sqrt{n-1}} \widehat{C}_p, \frac{\chi_{n-1, 1-\alpha/2}}{\sqrt{n-1}} \widehat{C}_p \right], \quad (2.10)$$

donde  $\chi_{n-1, a}^2$  es la raíz cuadrada del cuantil de orden  $a$  de la distribución Chi-cuadrado con  $n - 1$  grados de libertad. Es común sustituir este cuantil



por aproximaciones como las sugeridas por Fisher (1928) y Wilson y Hilferty (1931), y las cuales vienen dadas por:

$$\chi_{n,a} \cong \sqrt{n - \frac{1}{2}} + \frac{Z_a}{\sqrt{2}} = A_1, \quad (2.11)$$

y

$$\chi_{n,a} \cong n^{\frac{1}{2}} \left[ 1 - \frac{2}{9n} + Z_a \left( \frac{2}{9n} \right)^{\frac{1}{2}} \right] \frac{3}{2} = A_2, \quad (2.12)$$

donde  $Z_a$  es el cuantil de orden  $a$  de la distribución Normal estándar. Es decir, en lugar de trabajar con cuantiles de la distribución Chi-cuadrado, diversos autores han sugerido a través de estas aproximaciones utilizar cuantiles de la distribución Normal estándar. A continuación se analizan de forma exhaustiva estas aproximaciones.

En la Figura 2.12 se observan las diferencias en términos relativos ( $D1$  y  $D2$ ) entre ambas aproximaciones y el verdadero valor del cuantil de la distribución Chi-cuadrado para tamaños muestrales comprendidos entre 10 y 200, y los niveles de significación  $\alpha = \{0,01, 0,05, 0,1\}$ . En la Tabla 2.1 pueden consultarse los correspondientes órdenes de cuantiles usados en los intervalos de confianza, asociados a los comentados niveles de significación, y utilizados en la Figura 2.12. Las diferencias  $D1$  y  $D2$  se han calculado siguiendo las siguientes expresiones:

$$D1 = 100 \times \frac{\chi_{n;\alpha} - A_1}{A_1}$$

y

$$D2 = 100 \times \frac{\chi_{n;\alpha} - A_2}{A_2}.$$

A partir de estudio puede comprobarse que ambas aproximaciones  $A1$  y  $A2$  aportan resultados deseables para tamaños muestrales elevados, aunque la aproximación  $A1$  se aleja de forma significativa del verdadero valor del cuantil para tamaños muestrales pequeños.

Las expresiones analíticas de los intervalos de confianza, con nivel de significación  $1 - \alpha$ , definidos por la expresión (2.10) y basados en las aproximaciones (2.11) y (2.12) vienen dadas, respectivamente, por los intervalos  $[L_{A1}, U_{A1}]$  y  $[L_{A2}, U_{A2}]$ , donde

Tabla 2.1: Órdenes habituales de cuantiles usados en la construcción de intervalos de confianza con un nivel de significación  $\alpha$

Nivel de significación ( $\alpha$ )	$\alpha/2$	$1 - \frac{\alpha}{2}$
0.01	0.005	0.995
0.5	0.025	0.975
0.1	0.05	0.95

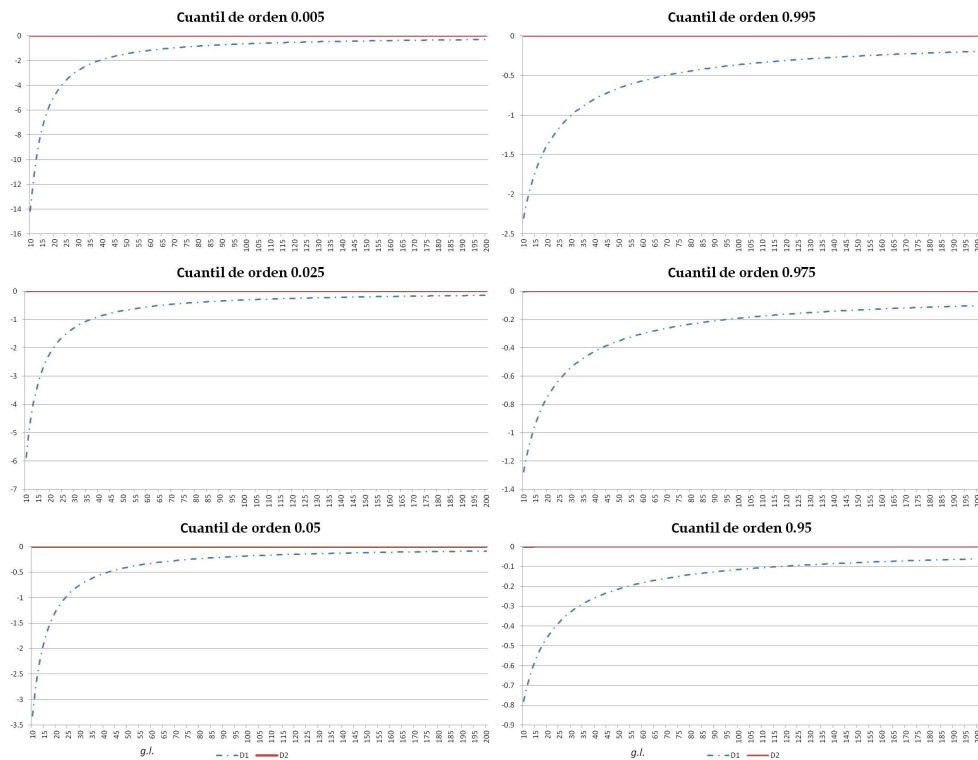


Figura 2.12: Valores de las diferencias  $D1$  y  $D2$  para cuantiles de órdenes usados en la construcción de los intervalos de confianza con niveles de significación  $\alpha = \{0,01, 0,05, 0,1\}$  y tamaños muestrales comprendidos entre 10 y 200.

$$L_{A1} = \frac{1}{\sqrt{n-1}} \widehat{C}_p \left( \sqrt{n - \frac{3}{2}} - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{2}} \right),$$

$$U_{A1} = \frac{1}{\sqrt{n-1}} \widehat{C}_p \left( \sqrt{n - \frac{3}{2}} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{2}} \right)$$

y

$$L_{A2} = \widehat{C}_p \left( 1 - \frac{2}{9(n-1)} - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{2}{9(n-1)}} \right),$$

$$U_{A2} = \widehat{C}_p \left( 1 - \frac{2}{9(n-1)} + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{2}{9(n-1)}} \right).$$

En este capítulo también se analiza el intervalo de confianza para el índice de capacidad *PCI* basado en el estimador  $\widehat{C}_{p.c4}$ . Este intervalo se define como  $[L_{c4}, U_{c4}]$ , y sus extremos se obtienen de forma similar, es decir, se sustituye  $\widehat{C}_p$  por  $\widehat{C}_{p.c4}$  en la ecuación (2.10). El intervalo de confianza para  $C_{pk}$  sugerido por Heavlin (1988) se define como  $[L_k, U_k]$ , donde

$$L_k = \widehat{C}_{pk} - Z_{1-\frac{\alpha}{2}} \left( \frac{n-1}{9n(n-3)} + \widehat{C}_{pk}^2 \frac{1}{2(n-3)} \left( 1 + \frac{6}{n-1} \right) \right)^{\frac{1}{2}}$$

y

$$U_k = \widehat{C}_{pk} + Z_{1-\frac{\alpha}{2}} \left( \frac{n-1}{9n(n-3)} + \widehat{C}_{pk}^2 \frac{1}{2(n-3)} \left( 1 + \frac{6}{n-1} \right) \right)^{\frac{1}{2}}$$

El intervalo de confianza para  $C_{pk}$  basado en  $\widehat{\sigma}_{c4}$ ,  $[L_{k.c4}, U_{k.v4}]$ , se define de forma similar. Otros estudios relacionados con la estimación mediante intervalos de confianza para  $C_{pk}$  pueden consultarse en Chou et al. (1990), Kushler y Hurley (1992), Franklin y Wasserman (1992), Guirguis y Rodriguez (1992), Dovich (1992), Nagata y Nagahata (1994), etc.

Los citados intervalos de confianza están basados en los Supuestos A1, A2 y A3. Sin embargo, existen situaciones donde estos supuestos no se cumplen. Por ejemplo, la teoría del muestreo de aceptación de lotes de productos en el campo del SQC asume la existencia de lotes finitos, es decir, el Supuesto

A2 no se cumple en esta situación. Una organización o empresa puede estar interesada en realizar un análisis de capacidad, y haber tomado los datos muestrales para el estudio de lotes finitos de productos. Si el tamaño del lote finito es elevado en comparación al tamaño muestral, podría usarse el intervalo de confianza basado en poblaciones infinitas, ya que está justificado su uso por la aproximación a una población infinita. Sin embargo, en otro caso el Supuesto A2 no se cumpliría, y podría ser interesante analizar diversos escenarios bajo estas circunstancias. Por este motivo, se ha investigado de forma numérica el efecto de esta aproximación para diferentes situaciones. Además, en la práctica podría darse el caso del incumplimiento de los Supuestos A1 y A2. En la Sección 2.2.3 se analiza numéricamente el efecto en los estimadores e intervalos de confianza para el índice de capacidad cuando no se verifican los Supuestos A1, A2 y A3. Para este propósito, se ha usado un estudio de simulación Monte Carlo donde se han considerado escenarios con diferentes características.

### 2.2.3. Estudio de simulación Monte Carlo

En la sección anterior se han descrito los estimadores e intervalos de confianza para los índices de capacidad  $PCI$  y  $PCI_k$ . Estos procedimientos de estimación pueden basarse en diferentes estimadores de la desviación típica del proceso, y por tanto pueden tener diferente comportamiento bajo ciertas situaciones, y especialmente en el caso de muestras con un tamaño pequeño. En primer lugar, en esta sección se investigan diversos escenarios mediante un estudio de simulación Monte Carlo, y el cual permitirá conocer las ventajas e inconvenientes de cada procedimiento de estimación y para cada una de las situaciones planteadas. Además, los intervalos de confianza habituales para los índices de capacidad  $PCI$  y  $PCI_k$  están basados los Supuestos A1, A2 y A3, pero tales supuestos pueden no cumplirse en la práctica. A continuación también se analiza numéricamente el efecto del incumplimiento de los supuestos básicos descritos. Finalmente, se comparan numéricamente los estimadores e intervalos de confianza basados en procesos centrados y no centrados, lo cual también permitirá analizar las posibles diferencias entre ambos tipos de procesos.

Destacamos que el presente estudio se da respuesta a los siguientes objetivos, los cuales se plantearon en la Sección 1.3:

- O1. Evaluar el comportamiento empírico del estimador de la desviación típica del proceso y su efecto en la estimación del  $PCI$ .

- O2. Evaluar el estimador del *PCI* en procesos no centrados.
- O3. Efecto en la estimación de la desviación típica y del *PCI* cuando el proceso está fuera de control.
- O4. Analizar el comportamiento empírico del estimador de la desviación típica y *PCI* en poblaciones infinitas y en poblaciones finitas.
- O5. Evaluar el comportamiento empírico del estimador de la desviación típica y del *PCI* en poblaciones “aproximadamente” infinitas (con una fracción de muestreo muy pequeña).
- O7. Estudio empírico del estimador de la desviación típica y del *PCI* en poblaciones que no presentan normalidad.
- O8. Evaluar el comportamiento empírico del estimador de la desviación típica y del *PCI* en una población basada en datos reales.

### Descripción de datos y escenarios usados en el estudio de simulación

En el estudio de simulación de esta sección se han considerado tres escenarios. En primer lugar, destacamos que los datos muestrales se seleccionan aleatoriamente de una distribución probabilística para cada repetición en la simulación. Esto implica que las variables son independientes e idénticamente distribuidas, es decir, las muestras son seleccionadas de poblaciones infinitas (se cumple el Supuesto A2). El segundo escenario consiste en seleccionar muestras usando muestreo aleatorio simple sin reemplazamiento de poblaciones finitas. En este caso,  $N$  valores son seleccionados aleatoriamente de la correspondiente distribución probabilística, y son considerados como “lotes” finitos desde donde se seleccionan muestras de tamaño  $n$ . Se consideran fracciones de muestreo ( $f = n/N$ ) elevadas con el objetivo de hacer más evidente las diferencias entre los casos de poblaciones finitas e infinitas. También se consideran fracciones de muestreo pequeñas con el objetivo de analizar numéricamente la aproximación de poblaciones finitas a poblaciones infinitas. Los tamaños de población considerados son  $N = \{500, 10000\}$ . Para poblaciones finitas se seleccionan muestras con varios tamaños  $n$  que van desde 10 a 500 en la población con tamaño  $N = 10000$ , lo que implica que las fracciones de muestreo oscilan entre 0,1 % y 5 % en esta situación. Finalmente, se considera el rango de valores que va desde 5 a 250 para el tamaño de muestra  $n$  y cuando el tamaño de la población es  $N = 500$ , lo que implica fracciones de muestreo con valores comprendidos entre el 1 % y el 50 %. Finalmente (el tercer escenario), también se

ha considerado el conjunto de datos reales “pistonrings”, comúnmente utilizados en el campo del SQC. Este conjunto de datos está disponible en el paquete “qcc” (Scrucca 2004) incluido en el software estadístico R (ver también Montgomery 2009). El conjunto de datos de pistonrings contiene valores relativos al diámetro interior de los anillos de pistones para un motor de automóvil y que son producidos por un proceso de forja. Siguiendo Harms y Duchesne (2006) se ha considerado este conjunto de datos reales como una población finita (con tamaño  $N = 125$ ), y a partir de la cual se seleccionarán las muestras. En este caso, se han considerado fracciones de muestreo con valores entre el 5 % y el 50 %.

Los datos aleatorios comentados con anterioridad se han generado de tres distribuciones probabilísticas continuas diferentes: Normal, Gamma y Uniforme. El objetivo es analizar el efecto del incumplimiento del Supuesto A1. Los valores aleatorios procedentes de la distribución Normal son obtenidos con media  $\mu = 10$ , y una desviación típica  $\sigma = 1$ , y que representan, respectivamente, la media y desviación típica de una característica de calidad de un determinado proceso. Para la distribución Gamma se han considerado los parámetros  $\alpha = 100$  y  $\theta = 1/10$  de esta distribución. Finalmente, los valores aleatorios de la distribución Uniforme son seleccionados entre 8.26795 y 11.73205. Como se puede apreciar, las distribuciones Gamma y Uniforme tienen la misma media ( $\mu = 10$ ) y la misma desviación típica ( $\sigma = 1$ ) que la distribución Normal. Varios contrastes de hipótesis no paramétricos indican que el conjunto de datos pistonrings se distribuyen según una distribución Normal. En este caso la media del proceso es  $\mu = 74,00288$  y la desviación típica del proceso viene dada por  $\sigma = 0,01291$ .

Como se observa, los escenarios previos comparan situaciones donde se cumplen los Supuestos A1 y A2 con situaciones donde estos supuestos no se verifican. Esta misma idea se sigue con el Supuesto A3. Con este propósito, primero se calculan los diferentes estimadores sin comprobar si el proceso está bajo control, es decir, se consideran todas las muestras seleccionadas (el Supuesto A3 no siempre se satisface en esta situación). Adicionalmente, también se obtienen los diferentes estimadores e intervalos de confianza cuando el proceso está bajo control, es decir, solo se consideran las muestras donde los estadísticos  $S$  y  $\bar{x}$  están dentro de los límites de control definidos por las expresiones (2.4) y (2.5), respectivamente. Tales límites de control se obtienen usando las constantes  $c = \{2,5,3\}$ . También se han eliminado las muestras que incumplen alguna de las reglas básicas que determinan si el proceso está o no bajo control. Destacamos que el porcentaje de muestras bajo control es bastante elevado, puesto que los datos son extraídos de la correspondiente distribución probabilística. Ésta puede ser la situación en la práctica, es decir, un

porcentaje elevado de muestras de una producción deberían estar bajo control.

Se han elegido límites de especificación ( $LIE$  y  $LSE$ ) tal que los índices de capacidad toman los valores  $C_p = C_{pk} = 1$ , es decir, se han considerado procesos centrados y no centrados, y donde  $k$  toma los valores  $k = \{0, 0,25, 0,5\}$ . En otras palabras, se han considerado los límites de especificación  $LIE = 7$  y  $LSE = 13$  cuando  $k = 0$  ( $C_p = 1; \mu_0 = 10$ ),  $LIE = 7$  y  $LSE = 15$  cuando  $k = 0,25$  ( $C_{pk} = 1; \mu_0 = 11$ ), y  $LIE = 7$  y  $LSE = 19$  cuando  $k = 0,5$  ( $C_{pk} = 1; \mu_0 = 13$ ) para las poblaciones generadas. Para el conjunto de datos reales (pistonrings) se han considerado los límites de especificación  $LIE = 73,96414$  y  $LSE = 74,04161$ , de forma que estos límites también proporcionan un índice de capacidad  $C_p = 1$  cuando  $k = 0$ .

### Medidas empíricas usadas en el estudio de simulación

Los diferentes estimadores e intervalos de confianza son analizados en términos de diferentes medidas empíricas, que son obtenidas después de  $B = 10000$  repeticiones en el estudio de simulación Monte Carlo. En particular, los estimadores se evalúan usando el sesgo relativo ( $RB$ ) y el error cuadrático medio relativo ( $RRMSE$ ). Destacamos que éstas son las medidas más comunes usadas para la comparación y el análisis de la precisión de estimadores (Chen y Sitter 1999, Deville y Särndal 1992, Muñoz et al. 2014, Rao et al. 1990, Silva y Skinner 1995). Por ejemplo, si se considera el estimador  $\hat{C}_{pk}$  del índice de capacidad  $C_{pk}$ , los correspondientes valores de  $RB$  y  $RRMSE$  se definen como:

$$RB_{PCI} = \frac{E[\hat{C}_{pk}] - C_{pk}}{C_{pk}} \quad (2.13)$$

y

$$RRMSE_{PCI} = \frac{\sqrt{MSE[\hat{C}_{pk}]}}{C_{pk}} \quad (2.14)$$

donde

$$E[\hat{C}_{pk}] = \frac{1}{B} \sum_{b=1}^B \hat{C}_{pk}(b)$$

es el valor esperado empírico de  $\widehat{C}_{pk}$  basado en las  $B$  simulaciones,  $\widehat{C}_{pk}(b)$  denota el valor de  $\widehat{C}_{pk}$  en la  $b$ -ésima repetición de la simulación, y

$$MSE[\widehat{C}_{pk}] = \frac{1}{B} \sum_{b=1}^B (\widehat{C}_{pk}(b) - C_{pk})^2$$

es el error cuadrático medio de  $\widehat{C}_{pk}$ . Los valores de  $RB$  y  $RRMSE$  descritos por las expresiones (2.13) y (2.14) se definen de forma similar en el caso de la estimación de la desviación típica  $\sigma$ , y se denotarán en este caso como  $RB_{DT}$  y  $RRMSE_{DT}$ , respectivamente.

Por su parte, los intervalos de confianza son evaluados usando su cobertura empírica para un nivel de confianza del 95 %. Por ejemplo, si  $L_k$  y  $U_k$  son el límite inferior y superior del intervalo de confianza para  $C_{pk}$ , la cobertura empírica correspondiente se define como

$$Coverage_{PCI} = \frac{1}{B} \sum_{b=1}^B I_{[L_k, U_k]}(b),$$

donde  $I_{[L_k, U_k]}(b)$  es una variable indicadora evaluada en la  $b$ -ésima simulación, y la cual toma el valor 1 si el parámetro  $C_{pk}$  está dentro del intervalo  $[L_k, U_k]$ , y el valor 0 en otro caso. Destacamos que es deseable que la cobertura empírica tome un valor cercano al nivel de confianza nominal (95 %). Los estimadores e intervalos de confianza se obtienen usando  $\widehat{\sigma}$  y  $\widehat{\sigma}_{c4}$ .

## Resultados derivados del estudio de simulación

En primer lugar, los estimadores  $\widehat{\sigma}$  y  $\widehat{\sigma}_{c4}$  se comparan en la Tabla 2.2 (objetivo O1). Se espera observar diferencias entre ellos cuando los tamaños muestrales son pequeños.

En las Figuras 2.13, 2.14 y 2.15 se consideran datos seleccionados de una distribución Normal, y para el problema de la estimación de  $C_{pk} = 1$ , con  $k = 0, 0,25, 0,5$ , respectivamente. En estas figuras se asume el caso de una población finita con un tamaño de  $N = 500$ . Adicionalmente, las Figuras 2.13, 2.14 y 2.15 incluyen las medidas empíricas cuando todas las muestras están bajo control, usando la constante  $c = 2,5$  para estos límites.



Tabla 2.2: Valores del Sesgo Relativo ( $RB(\%)$ ) y Error Cuadrático Medio Relativo ( $RRMSE(\%)$ ) para los estimadores de  $\sigma = 1$  (SD) y  $C_p = 1$  (PCI), y cobertura empírica ( $Coverage(\%)$ ) de intervalos de confianza para  $C_p = 1$  (PCI). Datos muestrales seleccionados de una distribución normal con  $\mu = 10$  y  $\sigma = 1$ . Para la población finita,  $N = 500$  datos son seleccionados de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.

Población	$n$	$RB_{SD}$		$RRMSE_{SD}$		$RB_{PCI}$		$RRMSE_{PCI}$		$Coverage_{PCI}$	
		$\hat{\sigma}$	$\hat{\sigma}_{c4}$	$\hat{\sigma}$	$\hat{\sigma}_{c4}$	$\hat{C}_p$	$\hat{C}_{p.c4}$	$\hat{C}_p$	$\hat{C}_{p.c4}$	$\hat{C}_p$	$\hat{C}_{p.c4}$
Infinita	5	-6.4	-0.8	34.0	35.1	25.2	18.0	67.3	31.3	95.5	94.3
	10	-2.8	-0.1	23.0	23.3	9.3	6.4	30.9	29.3	95.3	94.8
	15	-1.8	-0.1	18.4	18.6	5.7	3.9	22.1	21.4	95.4	95.1
	20	-1.3	0.0	16.0	16.1	4.1	2.8	18.3	17.8	95.3	95.0
	25	-0.9	-0.1	14.2	14.3	3.1	2.1	15.7	15.4	95.4	95.0
	50	-0.5	0.0	10.1	10.1	1.5	1.0	10.6	10.5	95.0	94.9
Finita ( $N = 500$ )	5	-10.4	-5.6	35.4	35.5	32.5	25.1	75.3	68.4	94.8	94.6
	10	-4.9	-2.4	24.1	24.2	12.5	9.5	34.0	32.1	94.0	93.9
	15	-5.2	-3.6	18.3	18.2	9.5	7.6	24.1	23.1	95.5	95.6
	20	-5.2	-4.0	16.8	16.6	8.6	7.3	21.1	20.3	94.1	94.4
	25	3.0	4.0	13.9	14.2	-1.1	-2.1	13.8	13.8	95.8	95.3
	50	-0.0	0.5	9.2	9.3	0.9	0.4	9.5	9.5	97.0	96.8

En segundo lugar, el efecto de diferentes valores de  $k$  en la estimación del índice de capacidad pueden observarse a través de las Figuras 2.13, 2.14 y 2.15 (objetivo O2).

En tercer lugar, destacamos que cada gráfico contiene resultados cuando: (i) se consideran todas las muestras; (ii) solo se consideran muestras bajo control. Esta cuestión permite analizar los efectos en estimadores e intervalos de confianza cuando el Supuesto A3 no se cumple en todos los casos (objetivo O3). El caso  $c = 3$  se muestra en las Figuras 2.16, 2.17 y 2.18. Destacamos que en esta situación, las diferencias entre los escenarios (i) y (ii) son pequeñas.

Adicionalmente, cada gráfico de las Figuras 2.13, 2.14 y 2.15 contienen ambos casos de poblaciones finitas ( $N = 500$ ) e infinitas, y cada una puede ser utilizada para analizar el Supuesto A2 (objetivo O4). De forma similar (objetivo O5), también se considera el caso de una población finita con un tamaño poblacional elevado ( $N = 10000$ ) y un tamaño muestral pequeño (Figuras 2.19 – 2.24). La aproximación a una población infinita claramente se sostiene en esta situación, y se comparan con los escenarios explicados anteriormente. Los parámetros usados en las Figuras 2.19, 2.20 y 2.21 son los mismos que en la Figuras 2.13, 2.14 y 2.15, respectivamente, es decir,  $C_{pk} = 1$ , con  $k = \{0, 0,25, 0,5\}$  y  $c = 2,5$  para los límites de control. En las Figuras 2.23, 2.23 y 2.24 se muestran los resultados derivados de escenarios donde se asumen los mismos valores para los parámetros a excepción de  $c = 3$ .

En sexto lugar, el efecto del incumplimiento del Supuesto A1 puede ser analizado a través de las Figuras 2.25 y 2.26, las cuales están basadas en datos extraídos de distribuciones Gamma y Uniforme (objetivo O7). El objetivo de las Figuras 2.25 y 2.26 es estimar el índice de capacidad  $C_{pk} = 1$ , con  $k = 0$ , y usando  $c = 3$  para los límites de control.

Finalmente, se usa la base de datos pistonrings para analizar los diferentes escenarios bajo esta situación real (objetivo O2). Los resultados correspondientes a este escenario se muestran en la Figura 2.27. Obviamente, solo se tiene el caso de una población finita. Sin embargo, puede cumplirse el Supuesto A2 en este caso si, por ejemplo, las muestras son seleccionados con reemplazamiento. Por esta razón, la Figura 2.27 contiene muestras seleccionadas con y sin reemplazamiento. Se considera  $C_{pk} = 1$ , con  $k = 0$ , usando  $c = 2,5$  para la simulación basada en la base de datos pistonrings.

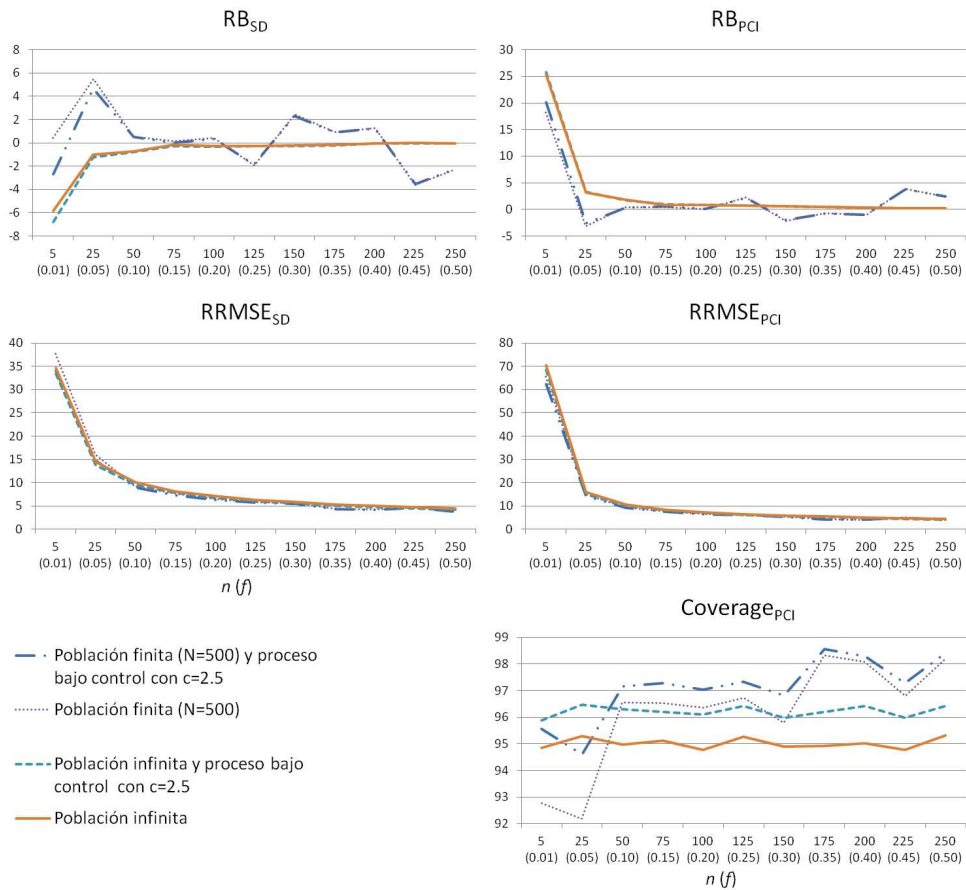


Figura 2.13: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_p = C_{pk} = 1$  ( $PCI$ ), y  $Coverage$  (%) de los intervalos de confianza para  $C_p = C_{pk} = 1$  ( $PCI$ ). Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 500$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.

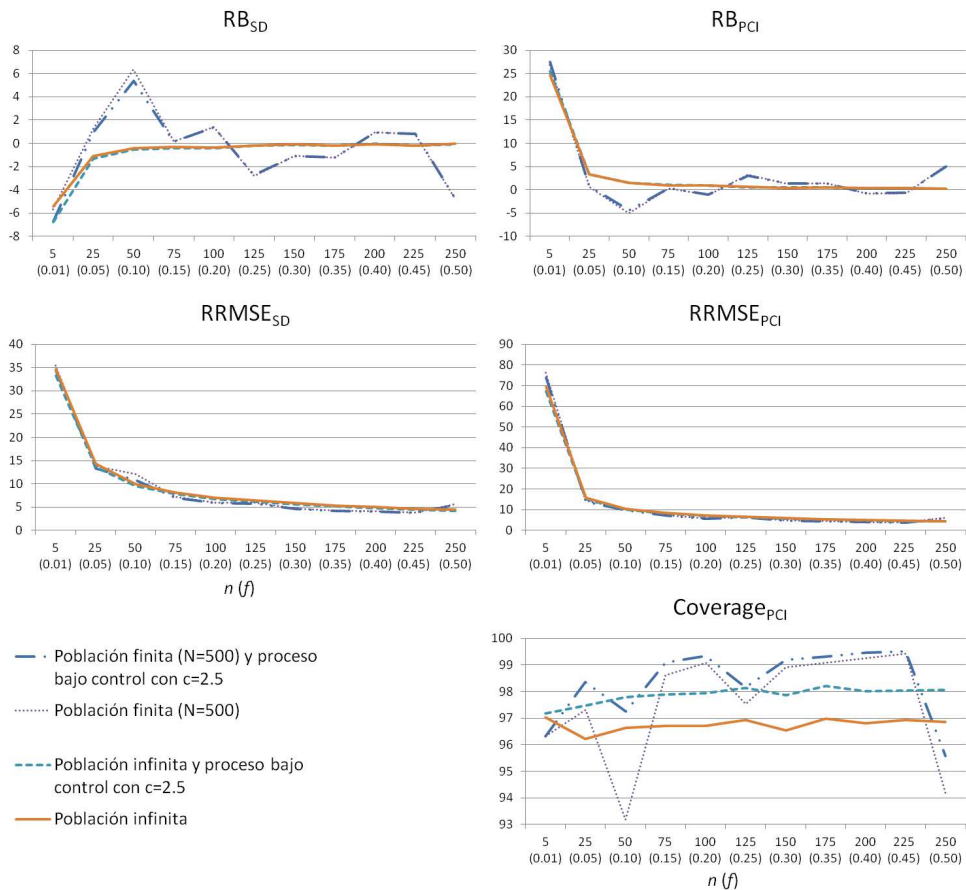


Figura 2.14: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,25$ , y  $Coverage$  (%) de los intervalos de confianza para  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,25$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 500$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.

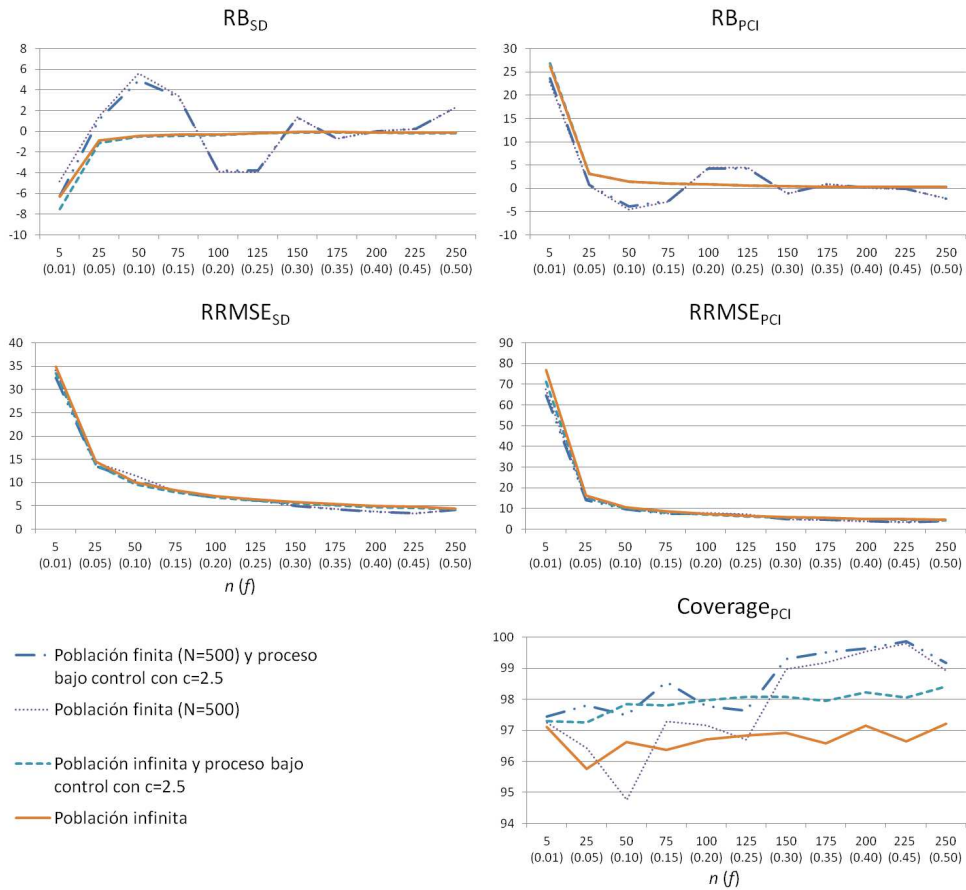


Figura 2.15: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,5$ , y  $Coverage$  (%) de los intervalos de confianza para  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,5$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 500$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.

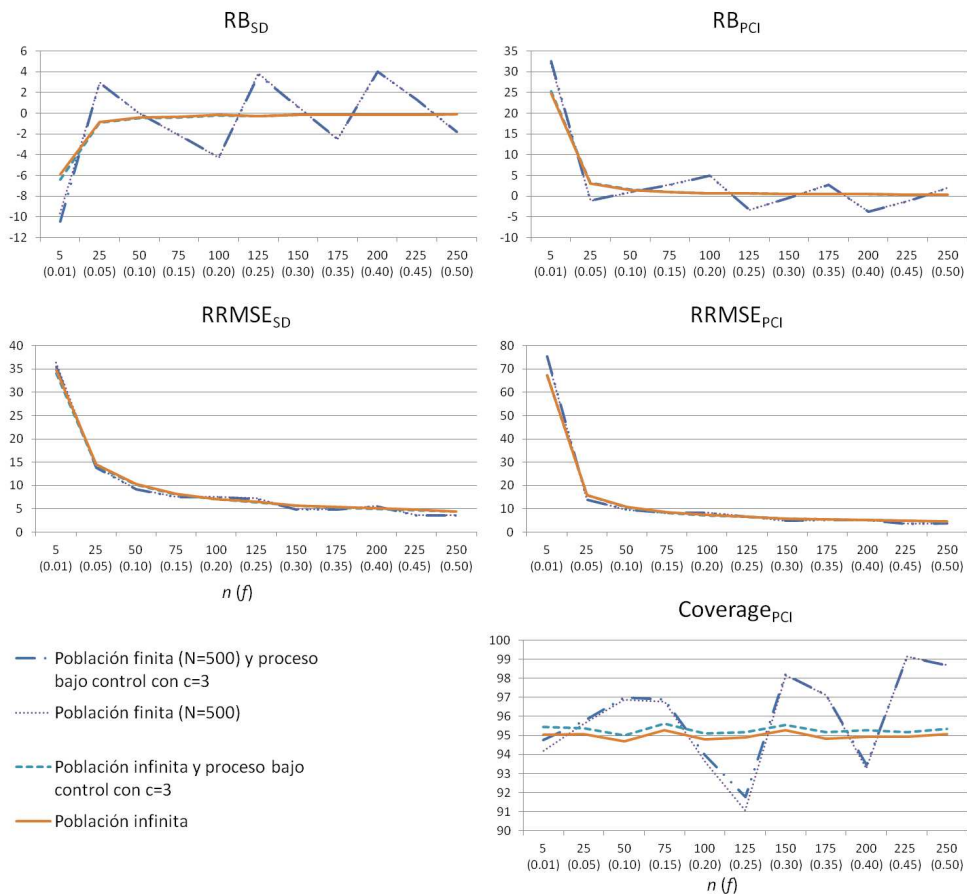


Figura 2.16: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_p = C_{pk} = 1$  ( $PCI$ ), y  $Coverage$  (%) de los intervalos de confianza para  $C_p = C_{pk} = 1$  ( $PCI$ ). Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 500$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.

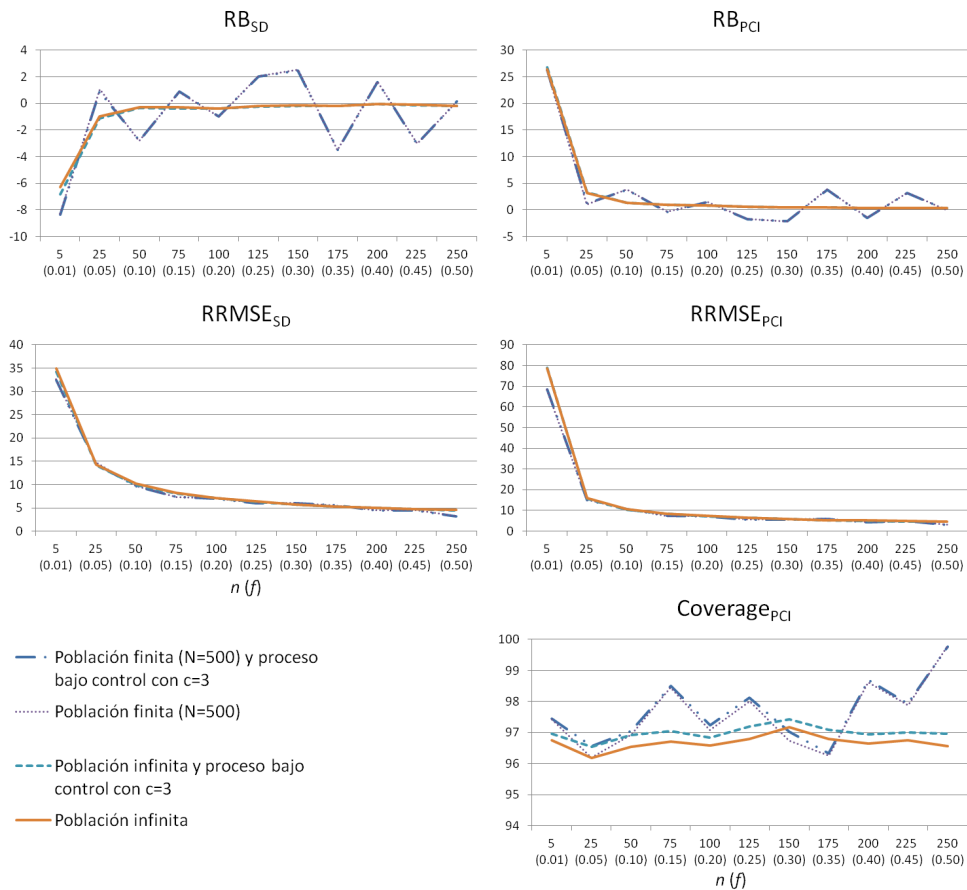


Figura 2.17: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,25$ , y  $Coverage$  (%) de los intervalos de confianza para  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,25$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 500$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.

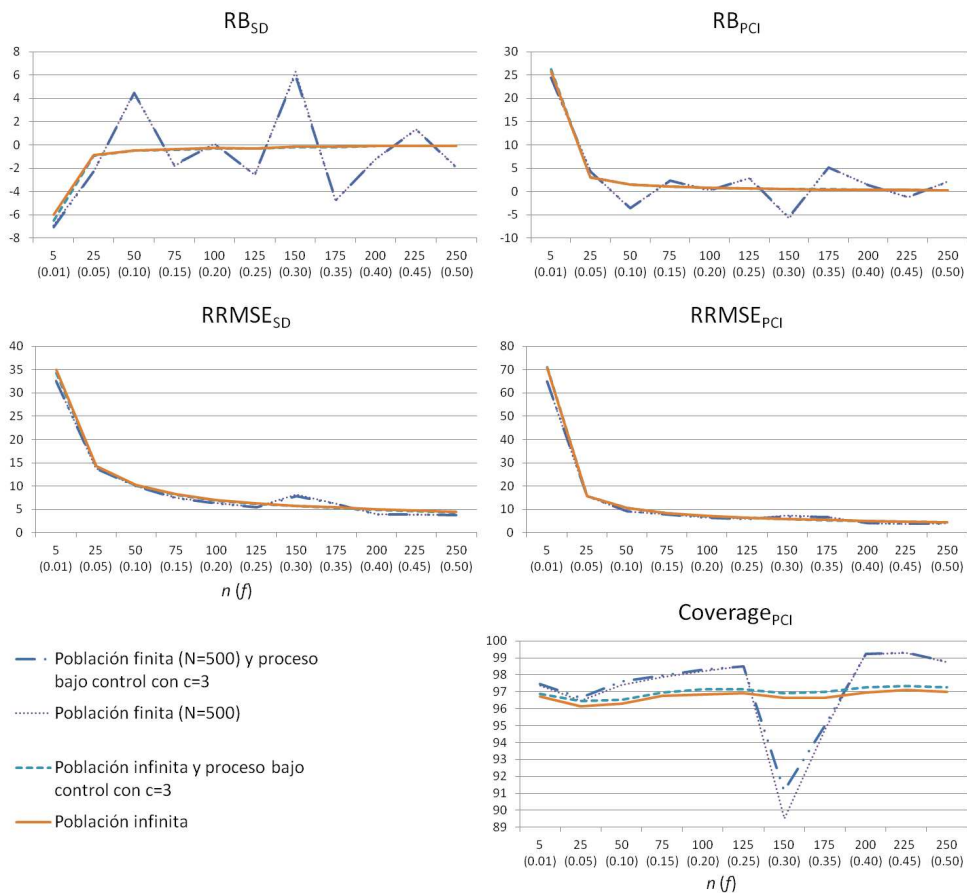


Figura 2.18: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,5$ , y  $Coverage$  (%) de los intervalos de confianza para  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,5$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 500$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.



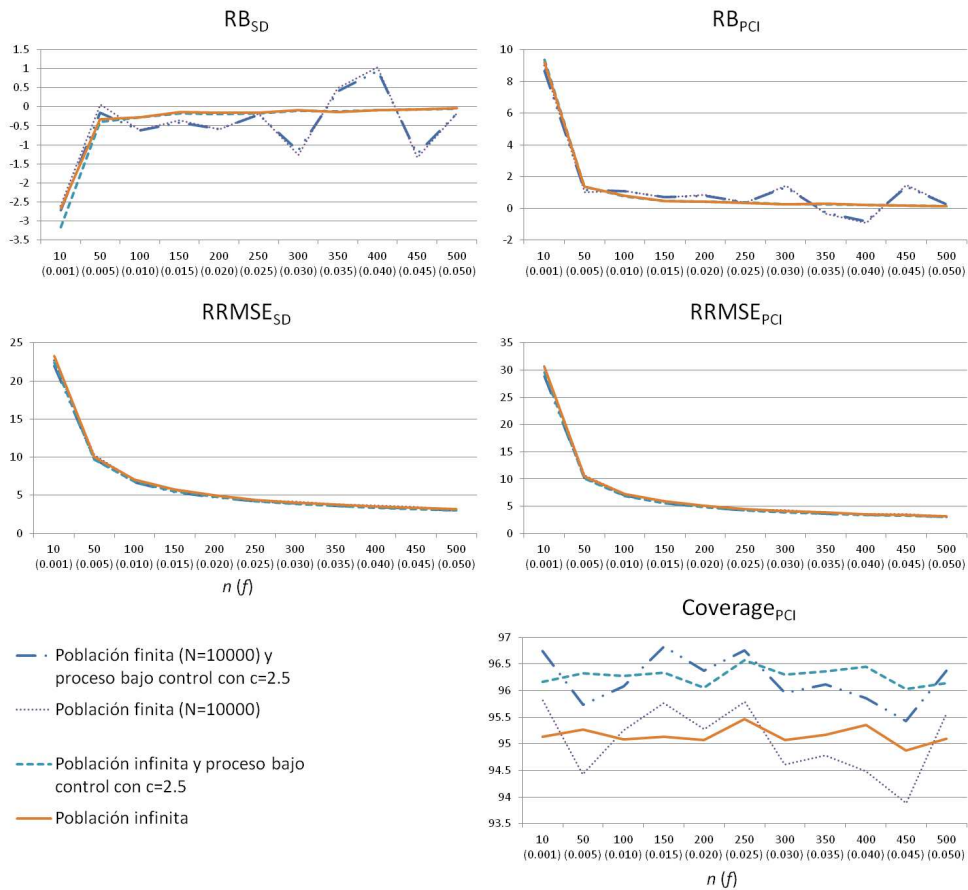


Figura 2.19: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_p = C_{pk} = 1$  ( $PCI$ ), y  $Coverage$  (%) de los intervalos de confianza para  $C_p = C_{pk} = 1$  ( $PCI$ ). Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 10000$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.

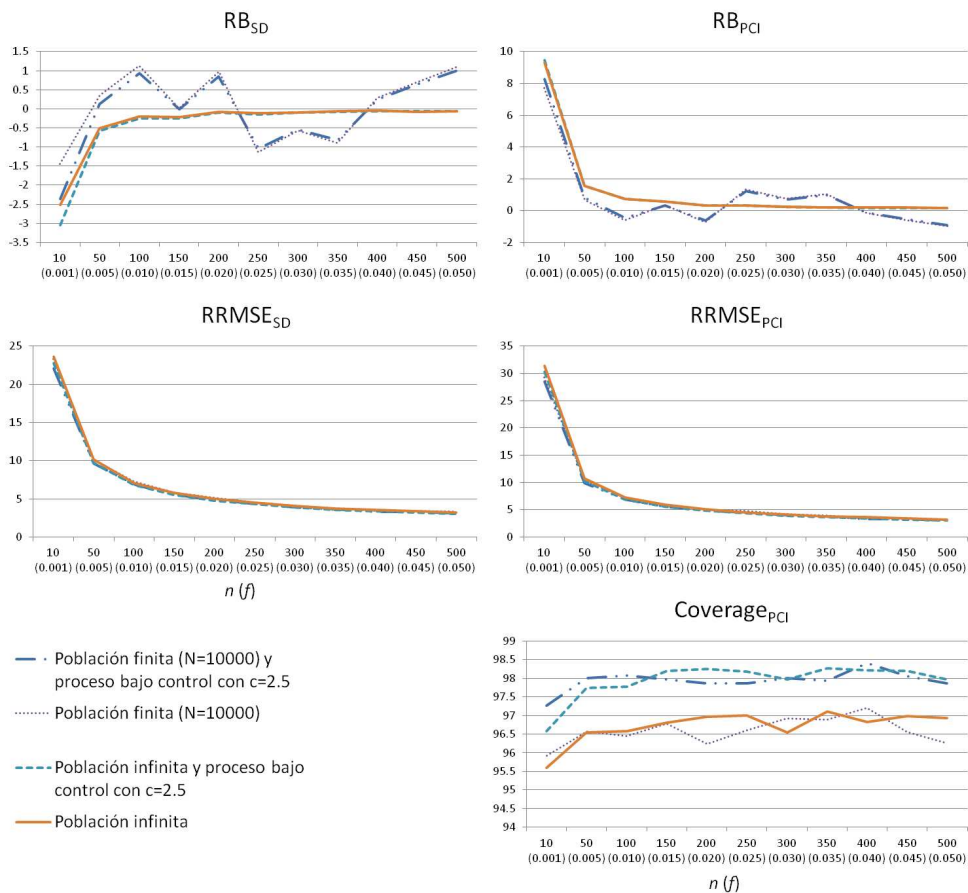


Figura 2.20: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,25$ , y  $Coverage$  (%) de los intervalos de confianza para  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,25$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 10000$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.

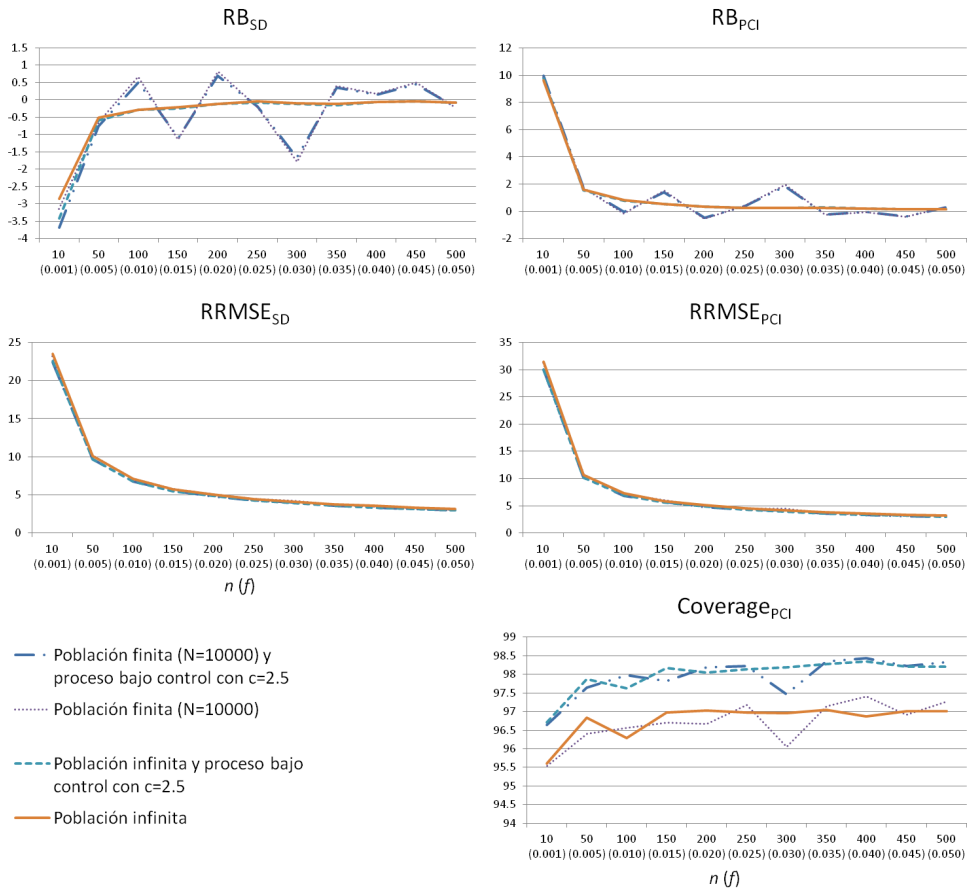


Figura 2.21: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,5$ , y  $Coverage$  (%) de los intervalos de confianza para  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,5$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 10000$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.

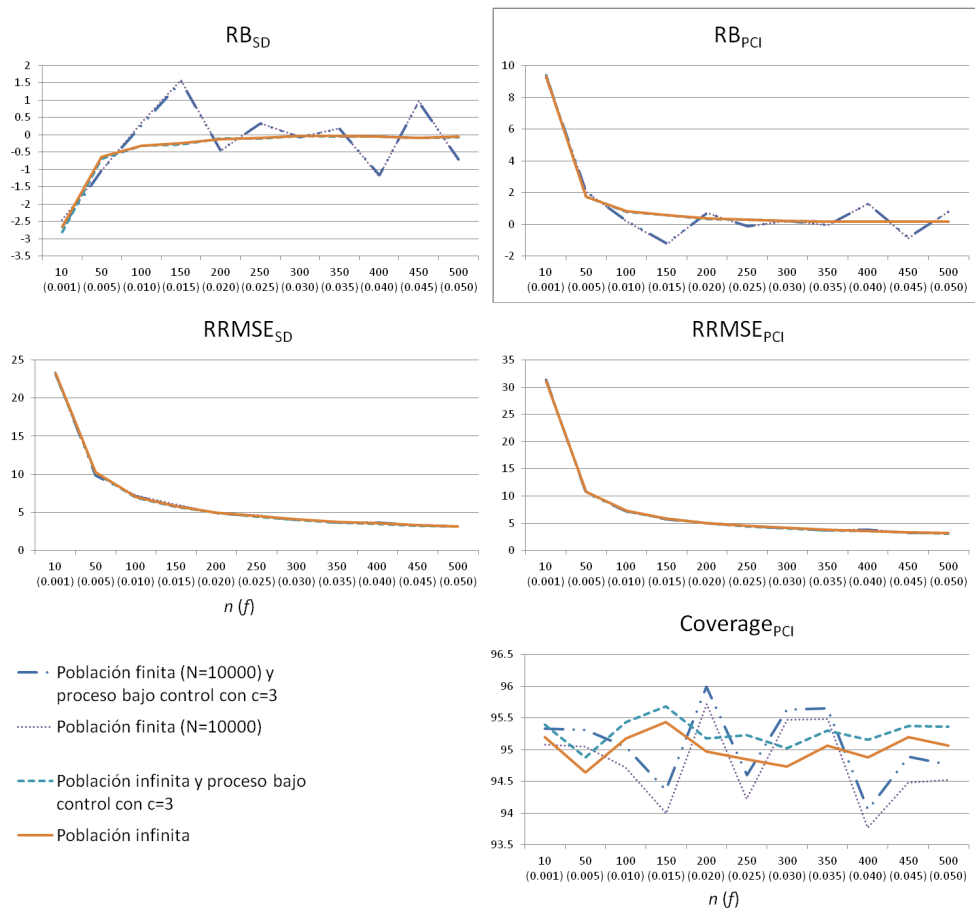


Figura 2.22: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_p = C_{pk} = 1$  ( $PCI$ ), y  $Coverage$  (%) de los intervalos de confianza para  $C_p = C_{pk} = 1$  ( $PCI$ ). Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 10000$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.

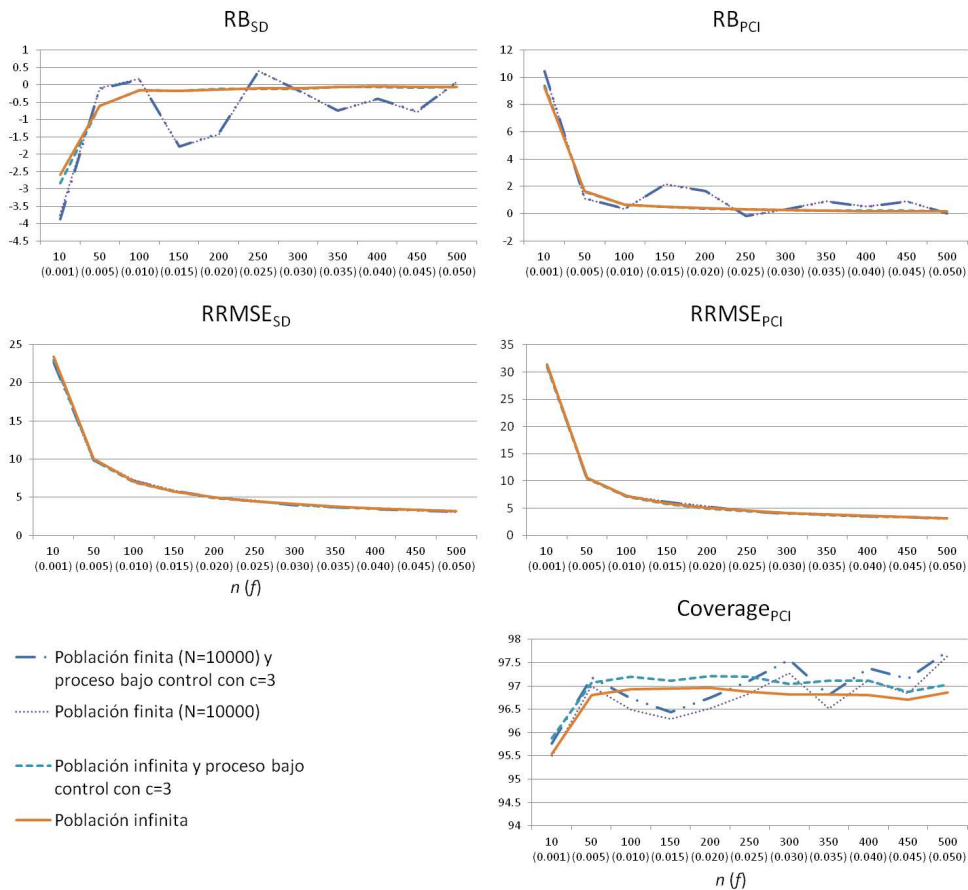


Figura 2.23: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,25$ , y  $Coverage$  (%) de los intervalos de confianza para  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,25$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 10000$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.

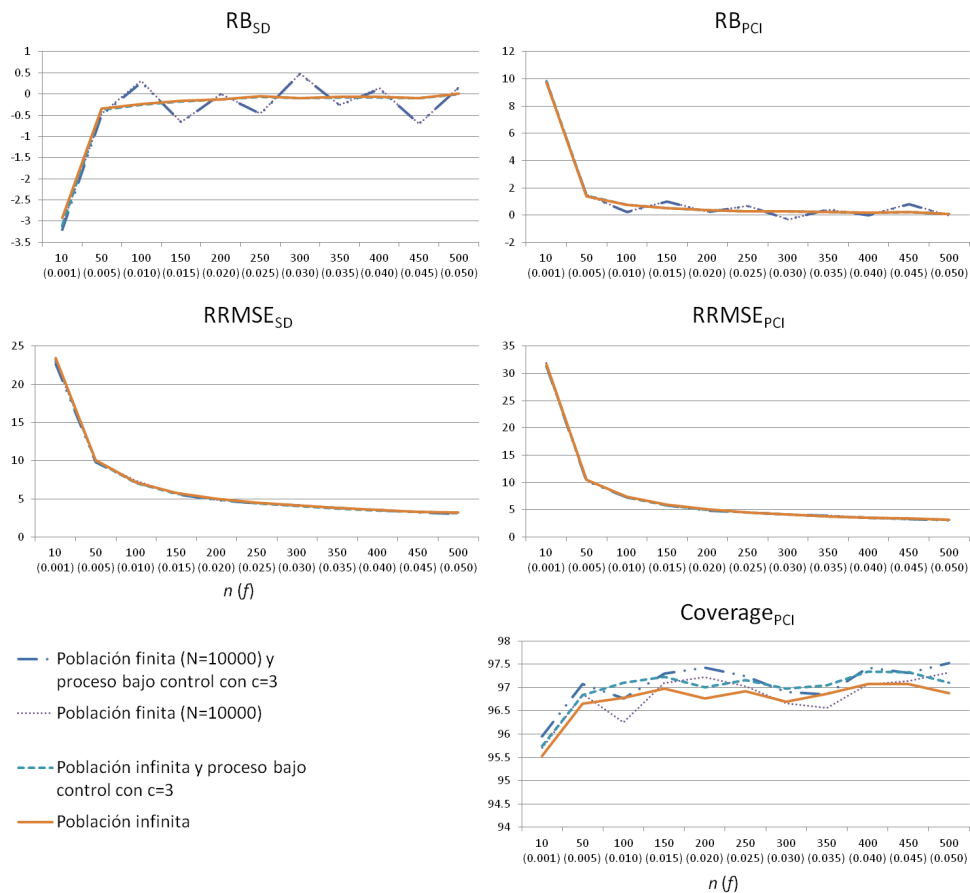


Figura 2.24: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,5$ , y  $Coverage$  (%) de los intervalos de confianza para  $C_{pk} = 1$  ( $PCI$ ), con  $k = 0,5$ . Se seleccionan los datos muestrales de una distribución Normal (población infinita). Para la población finita, se seleccionan  $N = 10000$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.

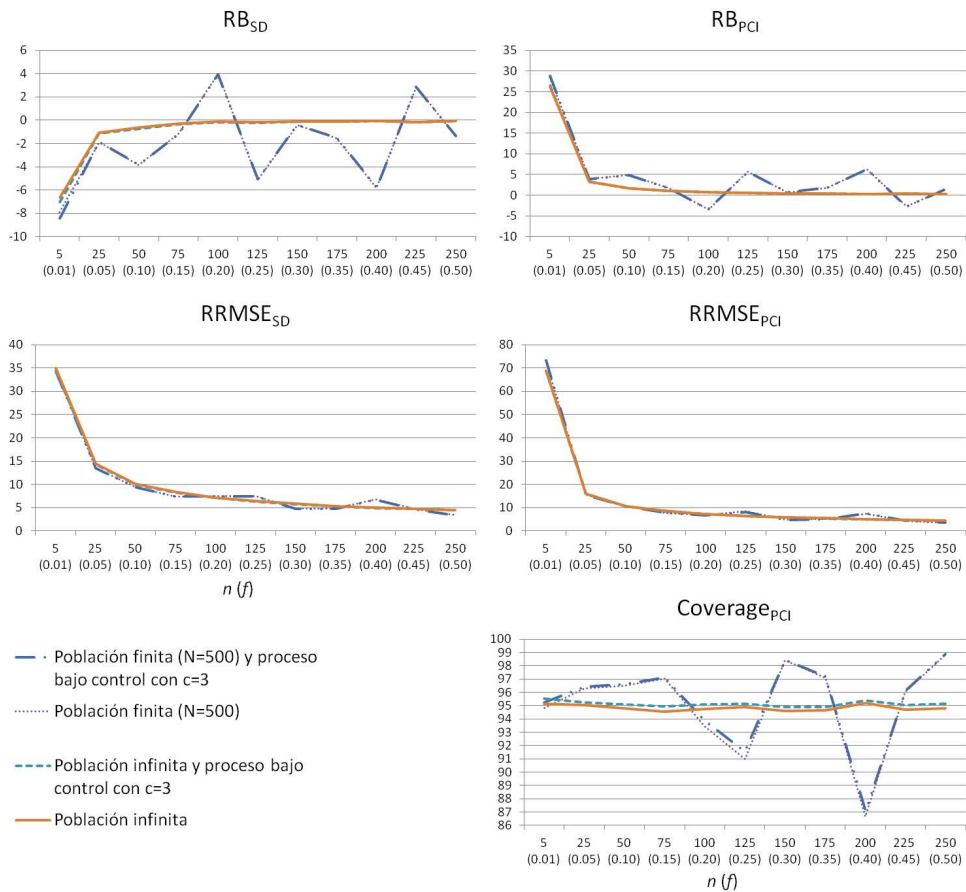


Figura 2.25: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_p = C_{pk} = 1$  ( $PCI$ ), y  $Coverage$  (%) de los intervalos de confianza para  $C_p = C_{pk} = 1$  ( $PCI$ ). Se seleccionan los datos muestrales de una distribución Gamma (población infinita). Para la población finita, se seleccionan  $N = 500$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.

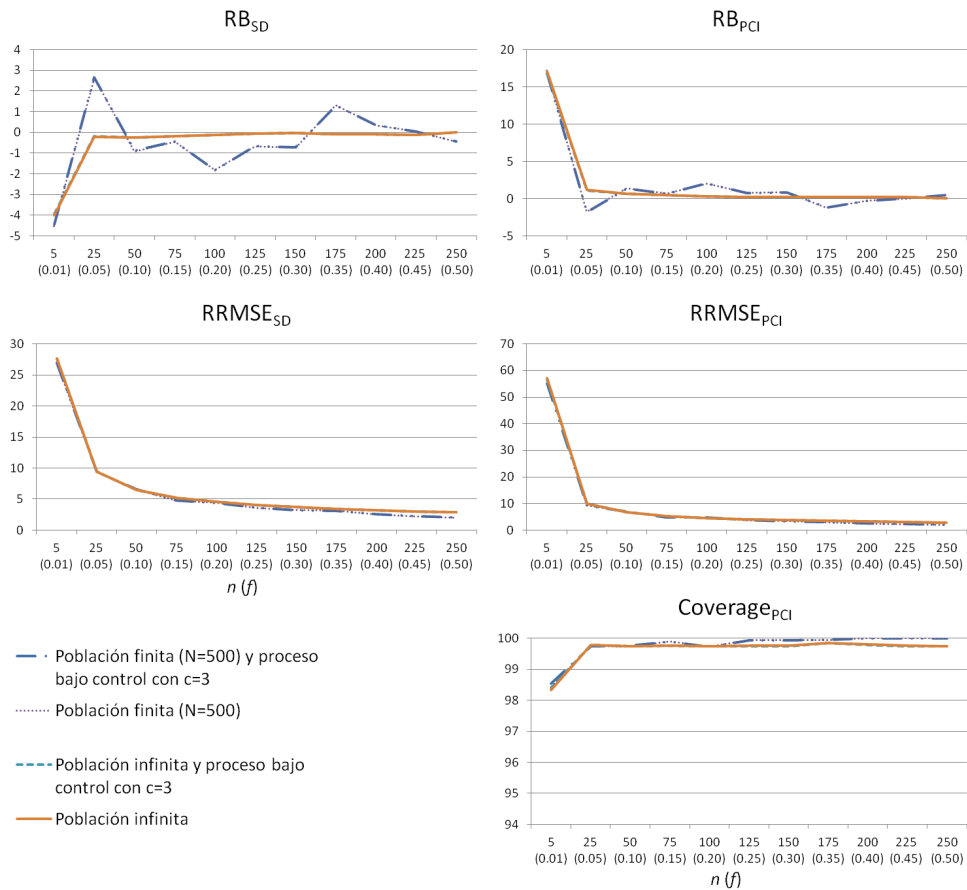


Figura 2.26: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 1$  ( $SD$ ) y  $C_p = C_{pk} = 1$  ( $PCI$ ), y  $Coverage$  (%) de los intervalos de confianza para  $C_p = C_{pk} = 1$  ( $PCI$ ). Se seleccionan los datos muestrales de una distribución Uniforme (población infinita). Para la población finita, se seleccionan  $N = 500$  datos de esta distribución y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple sin reemplazamiento.



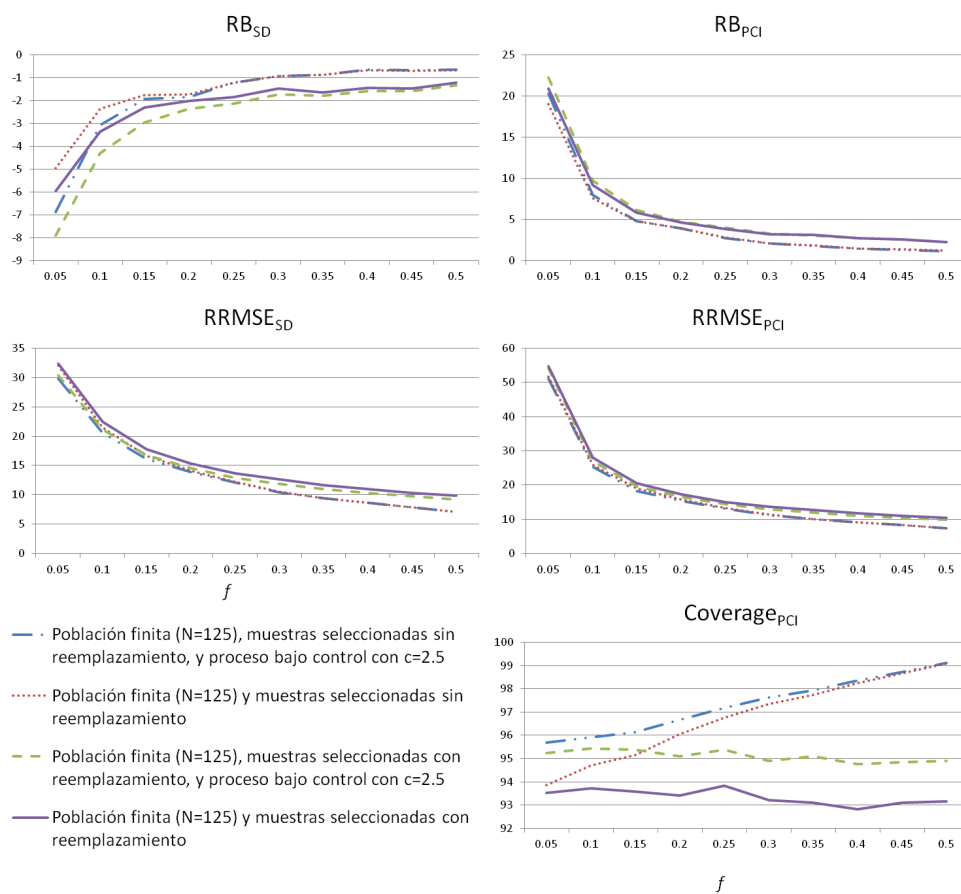


Figura 2.27: Valores de  $RB$  (%) y  $RRMSE$  (%) de los estimadores de  $\sigma = 0,01291$  ( $SD$ ) y  $C_p = C_{pk} = 1$  ( $PCI$ ), y  $Coverage$  (%) de los intervalos de confianza para  $C_p = C_{pk} = 1$  ( $PCI$ ). Se considera el conjunto de datos pistonrings como población, con  $N = 125$  y después se seleccionan las muestras de este conjunto de datos mediante muestreo aleatorio simple con y sin reemplazamiento.

## Discusión de los resultados empíricos

El comportamiento empírico de los estimadores  $\hat{\sigma}$  y  $\hat{\sigma}_{c4}$  se puede comparar en la Tabla 2.2. Como se esperaba, el estimador  $\hat{\sigma}$  tiene sesgos relativos ( $RB_{SD}$ ) mayores que el estimador  $\hat{\sigma}_{c4}$  cuando el tamaño muestral es pequeño, y ambos tienen un comportamiento similar a medida que el tamaño muestral incrementa.  $\hat{\sigma}_{c4}$  también presenta sesgos elevados en el caso de poblaciones finitas y tamaños muestrales pequeños (en torno a 5%). Si se analiza la precisión de los estimadores  $\hat{\sigma}$  y  $\hat{\sigma}_{c4}$  en términos de  $RRMSE_{SD}$ , se puede observar que ambos estimadores tienen un comportamiento similar, es decir,  $\hat{\sigma}$  es ligeramente más preciso que  $\hat{\sigma}_{c4}$  bajo varios tamaños de muestra. El estimador  $C_p$  basado en  $\hat{\sigma}$  ( $\hat{C}_p$ ) también tiene un mayor sesgo que el estimador  $\hat{C}_{p,c4}$ , y se observan diferencias más importantes cuando el tamaño muestral es menor que 20.  $\hat{C}_{p,c4}$  es generalmente más preciso que  $\hat{C}_p$  en términos de  $RRMSE_{PCI}$ . Generalmente se obtiene una cobertura empírica deseable, puesto que se obtienen valores cercanos al valor nominal del 95%. Se observan valores mayores del 95% cuando  $n = 50$  y la población es finita.

La Figura 2.13 muestra los valores de  $RB_{SD}$  y  $RRMSE_{SD}$  (en porcentaje) cuando la desviación típica del proceso es  $\sigma = 1$  y los datos muestrales han sido seleccionados de una distribución Normal. Se considera una población finita y una población infinita. Se observan sesgos relativos próximos al 0% en el caso de poblaciones infinitas y cuando el tamaño muestral es mayor que 25. Los valores de  $RB_{SD}$  basados en poblaciones finitas también son razonables, pero generalmente son valores con un rango comprendido entre el -4% y el 2%. Cuando  $n = 25$ , se obtienen sesgos relativos en torno al 6%. Los diferentes estimadores tienen un comportamiento similar en términos de  $RRMSE_{SD}$ . La Figura 2.13 también muestra los valores de  $RB_{PCI}$  y  $RRMSE_{PCI}$  (en porcentaje) cuando se estima el índice de capacidad  $C_p = 1$ . Las conclusiones obtenidas a partir de los sesgos relativos son similares a las conclusiones derivadas del caso de la estimación de  $\sigma = 1$ , es decir, los sesgos basados en poblaciones infinitas están cercanos al 0%, y con un rango de valores entre el -5% y el 5% en el caso de una población finita (para tamaños muestrales mayores que 25). Cuando las muestras tienen un tamaño menor que 25 se observan sesgos más elevados. Los resultados derivados de la Figura 2.13 también pueden ser utilizados para analizar el Supuesto A3, es decir, el hecho de que el proceso esté bajo control cuando se obtienen los estimadores y los intervalos de confianza. Se observa que no existen diferencias relevantes entre los estimadores en términos de sesgo ( $RB$ ) ni precisión ( $RRMSE$ ). Sin embargo, se aprecia que los intervalos de confianza obtenidos cuando el proceso está bajo control, y usando la constante  $c = 2,5$  para los límites de control, tienen una

cobertura empírica mayor que en el caso donde el Supuesto A3 no se cumple completamente. Destacamos que estas diferencias aparecen cuando  $c = 2,5$ , mientras que en la situación estándar (cuando  $c = 3$ ) se obtienen coberturas empíricas similares.

En la Figura 2.13 se muestra los resultados de un proceso centrado. Las Figuras 2.14 y 2.15 considera un proceso no centrado, donde  $C_{pk} = 1$  con  $k = 0,25, 0,5$ , respectivamente. Se observa que el valor de  $k$  no tiene un impacto relevante en los sesgos y precisión de los estimadores, ya que los valores de  $RB$  y  $RRMSE$  de las Figuras 2.13, 2.14 y 2.15 son similares. Sin embargo, se observa que el valor de  $k$  puede tener impacto en la cobertura de los intervalos de confianza, puesto que tales coberturas empíricas son mayores del 95 % cuando  $k = 0,25, 0,5$ .

El tamaño de la población no es elevado en la Figura 2.13 ( $N = 500$ ). La Figura 2.19 tiene las mismas condiciones que la Figura 2.13, excepto en el hecho de que el tamaño de la población es  $N = 10000$  en la Figura 2.19. En este último caso la población es finita, aunque existe una muy buena aproximación al caso de población infinita. Puede observarse la calidad de la mencionada aproximación si se comparan los resultados derivados de poblaciones finitas e infinitas en la Figura 2.19, es decir, los estimadores proporcionan coberturas similares en esta situación. Sin embargo, se obtienen diferentes coberturas en la Figura 2.13. Se obtienen resultados similares al comparar las Figuras 2.20 y 2.21 con las Figuras 2.14 y 2.15, respectivamente.

Si se analiza el incumplimiento de la hipótesis de normalidad (Supuesto A1) usando la distribución Gamma y poblaciones infinitas (Figura 2.25), se observa que el impacto no es importante en comparación a los resultados obtenidos bajo normalidad (ver, por ejemplo, la Figura 2.13). Asumiendo poblaciones finitas, las coberturas pueden tener mayor variabilidad en el escenario de la distribución Gamma (véase la Figura 2.25). Si se analiza una situación más extrema (la distribución Uniforme; Figura 2.26), se observa que el impacto es más relevante en los intervalos de confianza, puesto que la cobertura empírica correspondiente está cercana al 100 %. Esta situación se presenta para ambas poblaciones, finitas e infinitas.

En el caso del conjunto de datos reales se obtienen conclusiones similares. En primer lugar, los sesgos relativos se acercan a 0 % a medida que se incrementa la fracción de muestreo. Los diferentes escenarios presentan valores de  $RB$  y  $RRMSE$  similares. Por lo que se refiere a la cobertura empírica, se observa un buen comportamiento cuando se cumple el Supuesto A2 (las muestras son seleccionadas con reemplazamiento). Sin embargo, la cobertura empírica tiene

un peor comportamiento a medida que la fracción de muestreo se incrementa, asumiendo el escenario de muestras seleccionadas sin reemplazamiento.

## 2.2.4. Conclusiones

A continuación se resumen las conclusiones más importantes obtenidas en la Sección 2.2.3.

En primer lugar, se han analizado dos estimadores diferentes para la desviación típica y el índice de capacidad. Como se esperaba, el estimador sesgado de  $\sigma$  y de  $C_p$  tiene un mayor sesgo cuando los tamaños muestrales son menores que 10, y, en otro caso, el rendimiento de los estimadores es similar. Además, éstos tienen un comportamiento similar en términos de precisión y cobertura empírica.

Adicionalmente, también se analizan procesos centrados y no centrados, es decir, se consideran diferentes valores de  $k$ . Los resultados indican que el valor de  $k$  no tiene un impacto relevante ni en el sesgo ni en la precisión de los estimadores de  $C_p$ , pero valores de  $k$  mayores que 0 pueden producir intervalos de confianza con mayores coberturas empíricas.

Los estimadores e intervalos de confianza para el índice de capacidad están basados en los Supuestos A1, A2 y A3 descritos en la Sección 2.2.1. El Supuesto A1 indica que la característica de calidad sigue una distribución Normal. Si se considera la distribución Gamma y poblaciones infinitas se obtienen coberturas deseables. En el caso extremo de una distribución Uniforme, se obtienen coberturas cercanas al 100 %, bajo poblaciones finitas e infinitas.

El Supuesto A2 establece que las variables asociadas a los valores muestrales son independientes e idénticamente distribuidas. Se ha cumplido este supuesto cuando los datos muestrales han sido extraídos de una población infinita, y no se ha cumplido cuando los datos muestrales fueron seleccionados de una población finita. El primer efecto del incumplimiento de este supuesto es la presencia de sesgos con una mayor variación, pero dentro de unos valores razonables. Este supuesto no tiene un impacto relevante en la precisión, es decir, en los valores de  $RRMSE$ . El segundo efecto puede observarse en las coberturas empíricas de los intervalos de confianza. Los intervalos de confianza para el índice de capacidad, basados en poblaciones finitas, pueden tener coberturas mayores que el nivel nominal requerido. Los efectos comentados no son observados en poblaciones finitas con un tamaño poblacional elevado.

El Supuesto A3 indica que el índice de capacidad del proceso debe ser calculado cuando el proceso está bajo control (ver Montgomery 2009 p.181, Evans y Lindsay 1999 p. 321, Jensen et al. 2006). Se han analizado dos situaciones diferentes: (i) usando solo muestras cuyos estadísticos están dentro de los límites de control y no se incumple ninguna regla básica de los gráficos de control (el proceso está bajo control); y (ii) considerando todas las muestras sin comprobar si el proceso está bajo control. Se han obtenido resultados similares en términos de sesgo y precisión. Sin embargo, ambas situaciones pueden proporcionar coberturas diferentes para el intervalo de confianza del índice de capacidad.

### 2.3. Extensión al caso de $m$ -muestras

Como se ha comentado con anterioridad, uno de los supuestos básicos (A3) establece que el análisis de capacidad se lleva a cabo cuando el proceso está bajo control. Por tanto, si la empresa no tiene la certeza de que su proceso se encuentra en esta situación, deberá utilizar los gráficos de control. Es muy común asumir que  $\sigma$  es desconocido, ya que es poco probable conocer este parámetro en la práctica. En particular, la mayoría de los gráficos de control se basan en la estimación de  $\sigma$  (véase Chakraborty 2008, Chen 1997, Jones et al. 2001, Montgomery 2009 p. 228). En esta situación, el índice de capacidad de un proceso también requiere la estimación de la verdadera desviación típica  $\sigma$  del proceso.

Los parámetros desconocidos relacionados con el proceso se pueden estimar, en este caso, mediante el uso de las  $m$  muestras que se utilizan para elaborar los gráficos de control, donde  $m > 1$ , y las cuales se seleccionan cuando se cree que el proceso se encuentra bajo control. También es bastante común suponer que las diversas muestras tienen el mismo tamaño  $n$  puesto que esta metodología facilita la interpretación de los gráficos de control y otros resultados en el contexto del SQC. Se debe tener en cuenta que las expresiones para el caso de muestras con diferentes tamaños pueden ser fácilmente derivadas de la literatura existente (véase, por ejemplo, Montgomery 2009 p. 255). También es común el uso de muestreo aleatorio simple sin reemplazamiento para el problema de la selección de las  $m$  muestras.

A modo de extensión del estudio empírico llevado a cabo en la Sección 2.2, en esta sección se realiza un análisis comparativo para el caso en el que los parámetros son estimados utilizando la información contenida en las  $m$  muestras. Asumiendo este escenario, a continuación se definen los estimadores

más comunes de la desviación típica del proceso y los índices de capacidad de un proceso.

### 2.3.1. Estimadores de la desviación típica y del índice de capacidad

Sea  $\sigma$  el verdadero valor de la desviación típica de un proceso de producción. Los diferentes estimadores de  $\sigma$  están basados en los valores  $x_{ij}$  recogidos de las  $m$  muestras, donde  $x_{ij}$  denota el valor observado de la característica de calidad para el  $j$ -ésimo producto, con  $j = 1, \dots, n$ , en la  $i$ -ésima muestra, con  $i = 1, \dots, m$ . Para cada muestra seleccionada, también se asume el cumplimiento de los Supuestos A1, A2 y A3 definidos en la Sección 2.2.1.

El primer estimador de  $\sigma$  se define como

$$\hat{\sigma}_R = \frac{\bar{R}}{d_2[n]} \quad (2.15)$$

donde

$$\bar{R} = \frac{1}{m} \sum_{i=1}^m R_i$$

es la media de los rangos muestrales definidos como

$$R_i = \text{máx}(x_{ij}) - \text{mín}(x_{ij}).$$

La función  $d_2[\cdot]$  usada en la ecuación (2.15) depende del tamaño de muestra  $n$ , y su definición puede verse en Chen (1997). El valor de  $d_2[n]$  para distintos tamaños muestrales  $n$  se han tabulado en numerosas referencias relacionadas con el SQC. Por ejemplo, algunos valores de  $d_2[n]$  pueden consultarse en el Apéndice VI de Montgomery (2009). Destacar que  $\hat{\sigma}_R$  es un estimador insesgado de  $\sigma$  (véase Duncan 1986, Luko 1996, Wheeler 1995 y Woodall y Montgomery 2000).

Woodall y Montgomery (2000) definen el estimador

$$\hat{\sigma}_{R2} = \frac{d_2[n]}{d_2^*[n]} \bar{R}, \quad (2.16)$$

donde

$$d_2^*[n] = \left( d_2^2[n] + \frac{d_3^2[n]}{m} \right)^{1/2}$$

y los valores de  $d_3[n]$ , para los distintos valores de  $n$ , también están en general tabulados en referencias sobre el SQC (véase, por ejemplo, el Apéndice VI de Montgomery 2009). Destacar que el estimador  $\hat{\sigma}_{R2}$  tiene la propiedad de minimizar el error cuadrático medio entre todos los estimadores de la forma  $c\bar{R}$ , con  $c > 0$ . Por ejemplo, el estimador  $\hat{\sigma}_R$  está incluido en esta clase de estimadores.

Un tercer estimador ( $\hat{\sigma}_S$ ) de la desviación típica del proceso se puede obtener mediante el uso de las desviaciones típicas de las  $m$ -muestras. El estimador  $\hat{\sigma}_S$  se define como

$$\hat{\sigma}_S = \frac{\bar{S}}{c_4[n]} \quad (2.17)$$

donde

$$\bar{S} = \frac{1}{m} \sum_{i=1}^m S_i$$

es la media de la desviaciones típicas de las  $m$ -muestras, las cuales vienen definidas por:

$$S_i = \left( \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \right)^{1/2},$$

y donde

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$$

es la media de las muestras para los valores  $x_{ij}$  en la  $i$ -ésima muestra.

El último estimador de  $\sigma$  considerado en este trabajo se basa en la desviación típica ponderada de las muestras. Este estimador viene dado por

$$\hat{\sigma}_p = \frac{S_p}{c_4[1 + m(n-1)]}, \quad (2.18)$$

donde

$$S_p = \left( \frac{1}{m} \sum_{i=1}^m S_i^2 \right)^{1/2}$$

es la desviación típica ponderada basada en las  $m$  muestras con el mismo tamaño  $n$ . Cabe destacar que  $\hat{\sigma}_S$  y  $\hat{\sigma}_p$  también son estimadores insesgados de  $\sigma$  (véase Montgomery 2009 p. 253).

Las siguientes expresiones son, respectivamente, los estimadores de  $C_p$  basados en los estimadores de (2.15), (2.16), (2.17) y (2.18) del parámetro  $\sigma$ :

$$\widehat{C}_{p.R} = \frac{USL - LSL}{6\widehat{\sigma}_R}, \quad (2.19)$$

$$\widehat{C}_{p.R2} = \frac{USL - LSL}{6\widehat{\sigma}_{R2}}, \quad (2.20)$$

$$\widehat{C}_{p.S} = \frac{USL - LSL}{6\widehat{\sigma}_S}, \quad (2.21)$$

y

$$\widehat{C}_{p.P} = \frac{USL - LSL}{6\widehat{\sigma}_p}. \quad (2.22)$$

Los diversos estimadores de  $C_{pk}$  pueden definirse fácilmente mediante el uso de las diversas definiciones de los estimadores de  $C_p$ . Por ejemplo, el estimador de  $C_{pk}$  basado en la desviación estándar ponderada se define como

$$\widehat{C}_{pk.P} = (1 - k)\widehat{C}_{p.P}, \quad (2.23)$$

donde el estimador  $\widehat{C}_{p.P}$  está definido en la ecuación (2.22). Se puede observar que el estimador  $\widehat{C}_{pk.P}$  definido por la ecuación (2.23) asume que la media del proceso  $\mu$  es conocida. Si la media del proceso  $\mu$  es desconocida,

$$\widehat{\mu} = \bar{\bar{x}} = \frac{1}{m} \sum_{i=1}^m \bar{x}_i$$

es el estimador que en general se utiliza para la estimación de este parámetro.

### 2.3.2. Estudio de simulación Monte Carlo

En esta sección se analiza, a través de un estudio de simulación Monte Carlo, el comportamiento empírico de las diferentes estimadores del índice de capacidad del proceso ( $C_p$ ) en el caso de la selección de  $m$ -muestras. Asumiendo diferentes escenarios, se comparan los distintos estimadores de  $C_p$  en términos de sesgo y eficiencia. Para alcanzar este objetivo se utilizan las mismas medidas



que en el estudio anterior, es decir, el sesgo relativo ( $RB$ ) y el error cuadrático medio relativo ( $RRMSE$ ). Por razones de comparación, se analiza también el comportamiento empírico de los diferentes estimadores de la desviación típica del proceso ( $\sigma$ ). Este tema puede ser importante, puesto que puede ayudar en la interpretación del estado del proceso. Cabe destacar que la presencia, por ejemplo, de un sesgo significativo en el estimador de  $C_p$  puede producir una visión equivocada del estado del proceso. Además, una estimación eficiente del índice de capacidad del proceso es esencial para obtener una buena evaluación del proceso. En esta sección, se analiza el sesgo empírico y la eficiencia empírica de los distintos estimadores usuales de  $C_p$  definidos en las ecuaciones (2.19), (2.20), (2.21) y (2.22), y bajo un abanico amplio de escenarios susceptibles de presentarse en la práctica para un determinado proceso.

Destacamos que el presente estudio se corresponde con el objetivo O9 (estudio comparativo de los principales estimadores de la desviación típica y el  $PCI$  usando  $m$  muestras), y el objetivo O10 (estudio comparativo de los principales estimadores de la desviación típica y el  $PCI$  usando  $m$  muestras en poblaciones que no presentan normalidad). Los mencionados objetivos han sido descritos en la Sección 1.3.

Este estudio de simulación está basado en  $B = 1000$  réplicas. En la primera ejecución de la simulación,  $m$  muestras con el mismo tamaño  $n$  se seleccionan de una distribución probabilística con una desviación típica  $\sigma$ . Estos valores pueden representar la característica de calidad de un determinado elemento dentro de un proceso de producción. Los valores de  $\sigma$  y de los límites de especificación ( $LIE$  y  $LSE$ ) se seleccionan de forma que se obtienen diferentes valores de  $C_p$ . Esta información se utiliza para obtener el verdadero índice de capacidad del proceso  $C_p$ , y los diversos estimadores de este parámetro definido por las ecuaciones (2.19), (2.20), (2.21) y (2.22). Este proceso se repite  $B = 1000$  veces. En este estudio se han considerado los valores  $m = \{10, 20, 30, 50, 100, 1000\}$ ,  $\sigma = \{1, 3\}$  y  $C_p = \{1, 1,5\}$ . Los tamaños de muestra  $n$  toman valores entre 3 y 25, en una secuencia con paso 2, y las muestras se seleccionan bajo muestreo aleatorio simple sin reemplazamiento.

Las distribuciones Normal, Gamma y Uniforme son las distribuciones probabilísticas utilizadas en este estudio. La distribución Normal se considera porque éste es el Supuesto A1. Las distribuciones Gamma y Uniforme se consideran para analizar el impacto en los distintos estimadores del índice de capacidad de proceso cuando se tienen en cuenta distribuciones alternativas (incumplimiento del Supuesto A1). Por último, también se han considerado procesos no centrados. En esta situación, los límites de especificación se seleccionan de forma que  $C_{pk} = \{1, 1,5\}$ .

En las Figuras 2.28, 2.30, 2.32 y 2.34 se pueden observar los valores de  $RB$  de los estimadores de  $C_p = 1$ , mientras que las Figuras 2.29, 2.31, 2.33 y 2.35 muestran los valores de  $RB$  de los estimadores de  $\sigma = 1$ . Suponiendo datos procedentes de la distribución Normal y el problema de la estimación de  $C_p$  (Figura 2.28), se observan sesgos elevados cuando  $n = \{3, 5\}$ . Los diferentes estimadores proporcionan sesgos pequeños cuando  $n > 5$  y  $m = 10$ , con valores de  $RB$  alrededor de 0,5% en esta situación. El comportamiento de los diferentes estimadores es similar cuando  $m$  es distinto de 10, pero los sesgos se acercan a 0 cuando se incrementa  $m$ . Los valores de  $RB$  del estimador  $\widehat{C}_{p,R2}$  son ligeramente más grandes que los valores de  $RB$  del resto de estimadores del índice de capacidad, especialmente para valores pequeños de  $m$ . Asumiendo distribución Normal (Figura 2.29) se observa que los estimadores de  $\sigma$  tienen un buen comportamiento empírico en términos de sesgo. El estimador  $\widehat{\sigma}_{R2}$  tiene valores de  $RB$  cercanos a -2% cuando  $n = 3$  y  $m = 10$ .

Se observa que los valores de  $RB$  basados en la distribución Gamma (Figuras 2.30 y 2.31) son ligeramente mayores que los valores de  $RB$  basados en la distribución Normal (Figuras 2.28 y 2.29). Por ejemplo, los valores de  $RB$  de los estimadores  $\widehat{C}_{p,R}$ ,  $\widehat{C}_{p,S}$  y  $\widehat{C}_{p,P}$  están sobre el 3,5% cuando  $n = 3$ ,  $m = 10$  y se ha usado la distribución Gamma (ver Figura 2.30), mientras que los valores correspondientes de  $RB$  basados en la distribución Normal se acercan al 2% (ver Figura 2.28) cuando  $n = 3$  y  $m = 10$ . Suponiendo la distribución Gamma, los estimadores de  $\sigma$  también tienen un buen comportamiento empírico en términos de sesgo, aunque el estimador  $\widehat{C}_{p,R2}$  presenta sesgos grandes cuando ambos valores de  $n$  y  $m$  son pequeños. Estos resultados indican que el impacto en los diferentes estimadores de  $C_p$  y  $\sigma$  no es relevante si se usa la distribución Gamma en lugar de la distribución Normal.

Una distribución extrema en comparación con la distribución Normal es la distribución Uniforme. La distribución Uniforme se caracteriza por el hecho de que todos los intervalos de la misma longitud en el recorrido de la distribución son igualmente probables. En las Figuras 2.32 y 2.33 se observa que los sesgos de los estimadores basados en los rangos de la muestra ( $\widehat{C}_{p,R}$ ,  $\widehat{C}_{p,R2}$ ,  $\widehat{\sigma}_R$  y  $\widehat{\sigma}_{R2}$ ) aumentan a medida que los tamaños de muestra  $n$  incrementan. Por lo tanto, se puede concluir que la distribución Uniforme tiene un impacto importante en el comportamiento de los estimadores basados en los rangos de la muestra. Para valores pequeños de  $n$ , los valores de  $RB$  de los estimadores basados en la desviación típica de la muestra ponderada son ligeramente menores que los valores de  $RB$  de los estimadores basados en las desviaciones típicas de la muestra, y todos ellos se aproximan a cero a medida que incrementa  $n$ .

Por último, se analizan los sesgos relativos cuando el proceso no está cen-

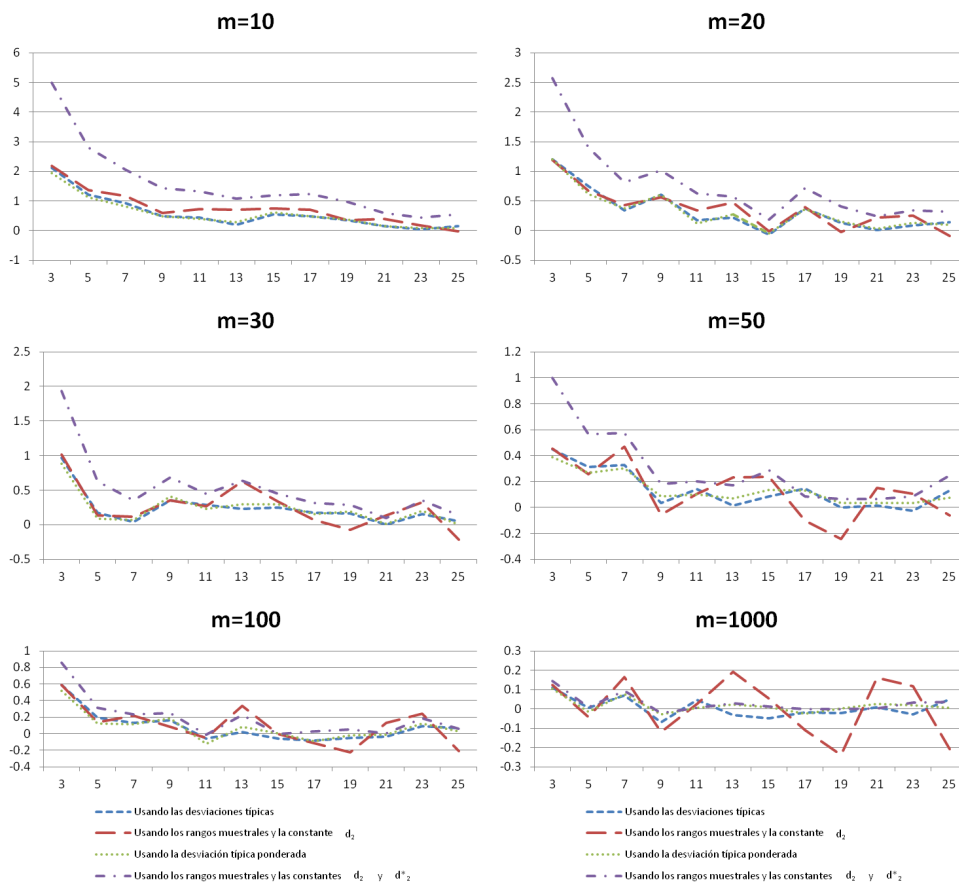


Figura 2.28: Valores de  $RB$  (%) de los estimadores  $\widehat{C}_{p.R}$  (usando los rangos muestrales y la constante  $d_2$ ),  $\widehat{C}_{p.R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\widehat{C}_{p.S}$  (usando la desviación típica muestral) y  $\widehat{C}_{p.P}$  (usando la desviación típica ponderada) de  $C_p = 1$ . Datos seleccionados de la distribución Normal con  $\sigma = 1$ .

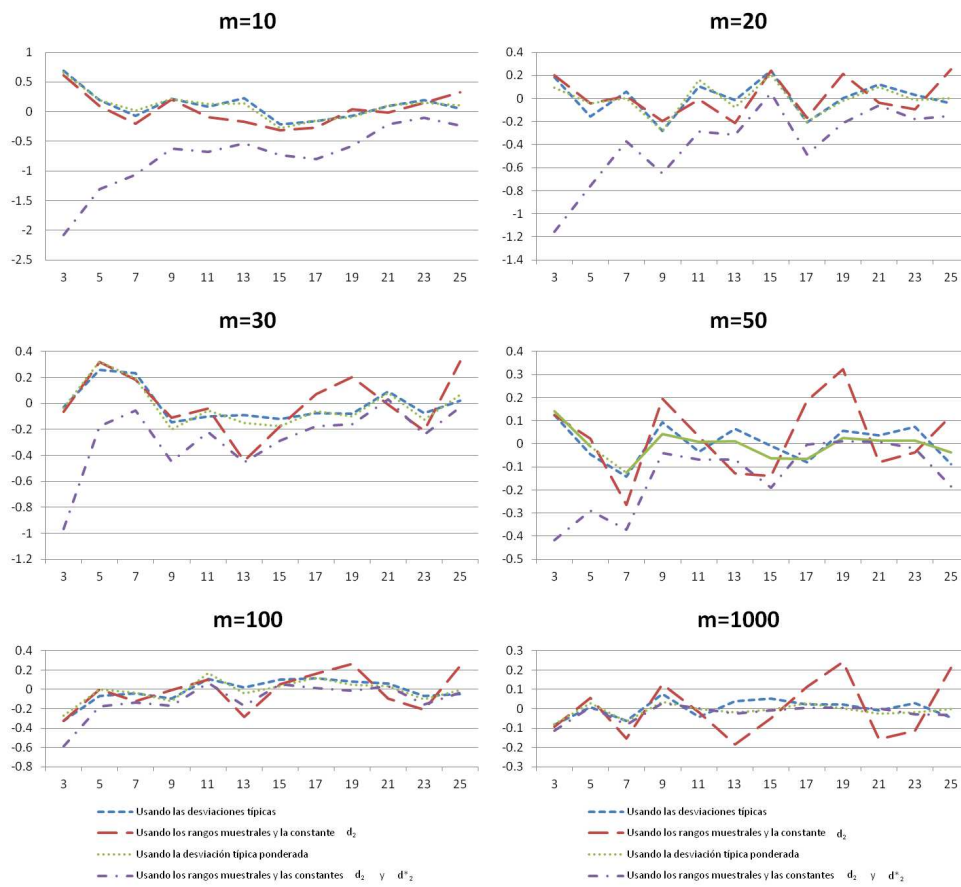


Figura 2.29: Valores de  $RB$  (%) de los estimadores  $\hat{\sigma}_R$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{\sigma}_{R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{\sigma}_S$  (usando la desviación típica muestral) y  $\hat{\sigma}_P$  (usando la desviación típica ponderada) de  $\sigma = 1$ . Datos seleccionados de la distribución Normal.

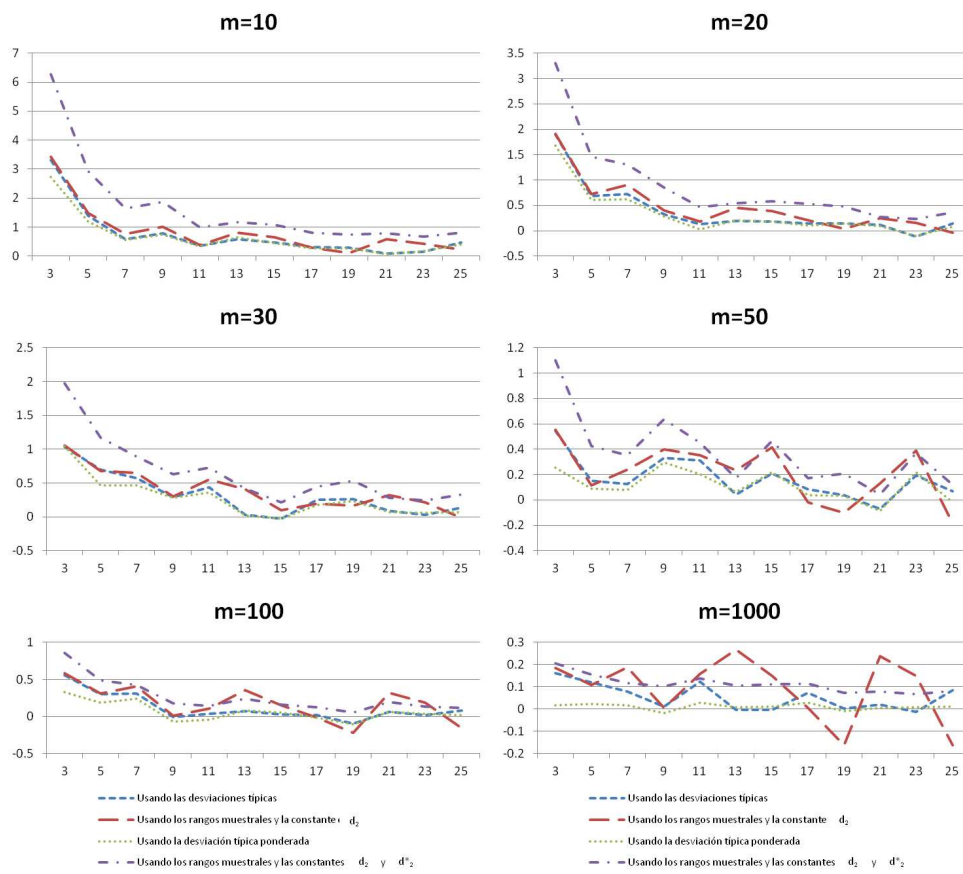


Figura 2.30: Valores de  $RB$  (%) de los estimadores  $\hat{C}_{p,R}$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{C}_{p,R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{C}_{p,S}$  (usando la desviación típica muestral) y  $\hat{C}_{p,P}$  (usando la desviación típica ponderada) de  $C_p = 1$ . Datos seleccionados de la distribución Gamma con  $\sigma = 1$ .

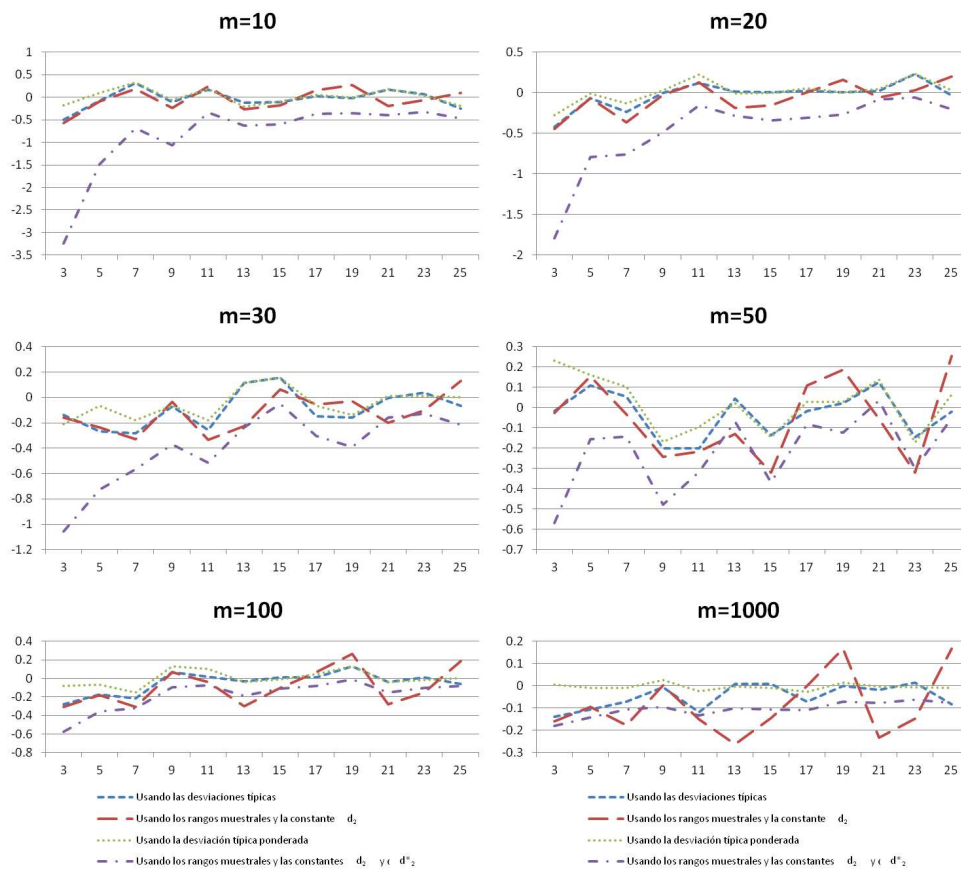


Figura 2.31: Valores de  $RB$  (%) de los estimadores  $\hat{\sigma}_R$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{\sigma}_{R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{\sigma}_S$  (usando la desviación típica muestral) y  $\hat{\sigma}_P$  (usando la desviación típica ponderada) de  $\sigma = 1$ . Datos seleccionados de la distribución Gamma.

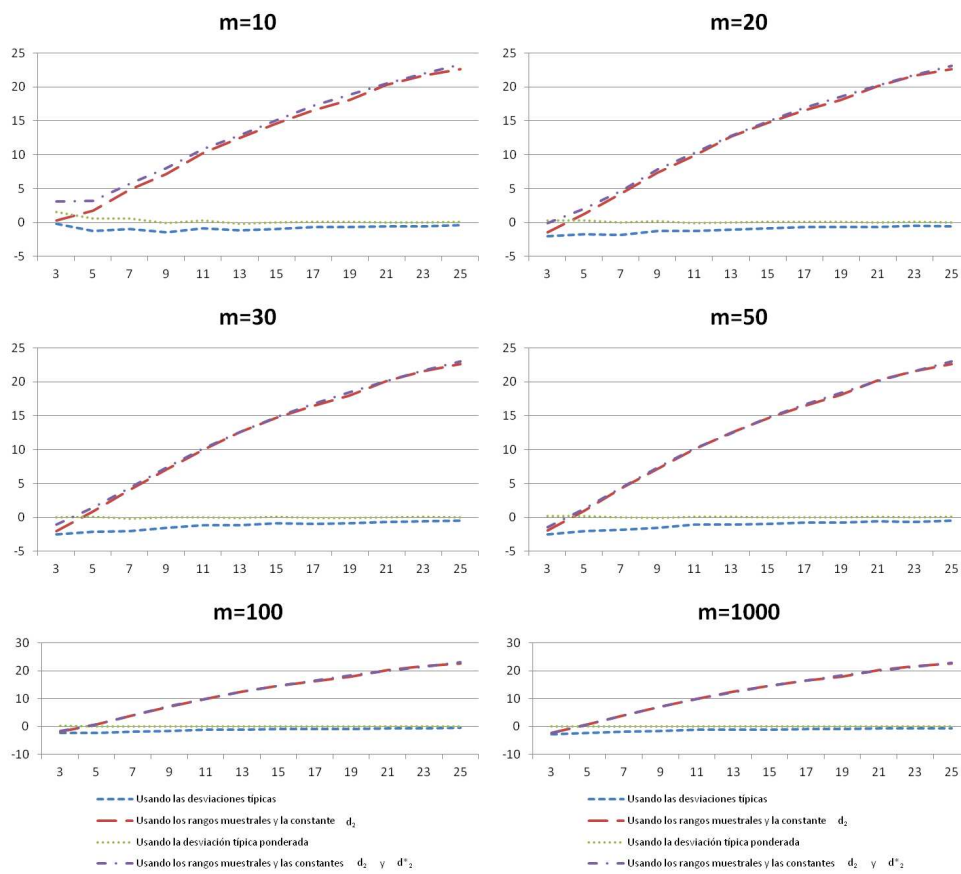


Figura 2.32: Valores de  $RB$  (%) de los estimadores  $\hat{C}_{p.R}$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{C}_{p.R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{C}_{p.S}$  (usando la desviación típica muestral) y  $\hat{C}_{p.P}$  (usando la desviación típica ponderada) de  $C_p = 1$ . Datos seleccionados de la distribución Uniforme con  $\sigma = 1$ .

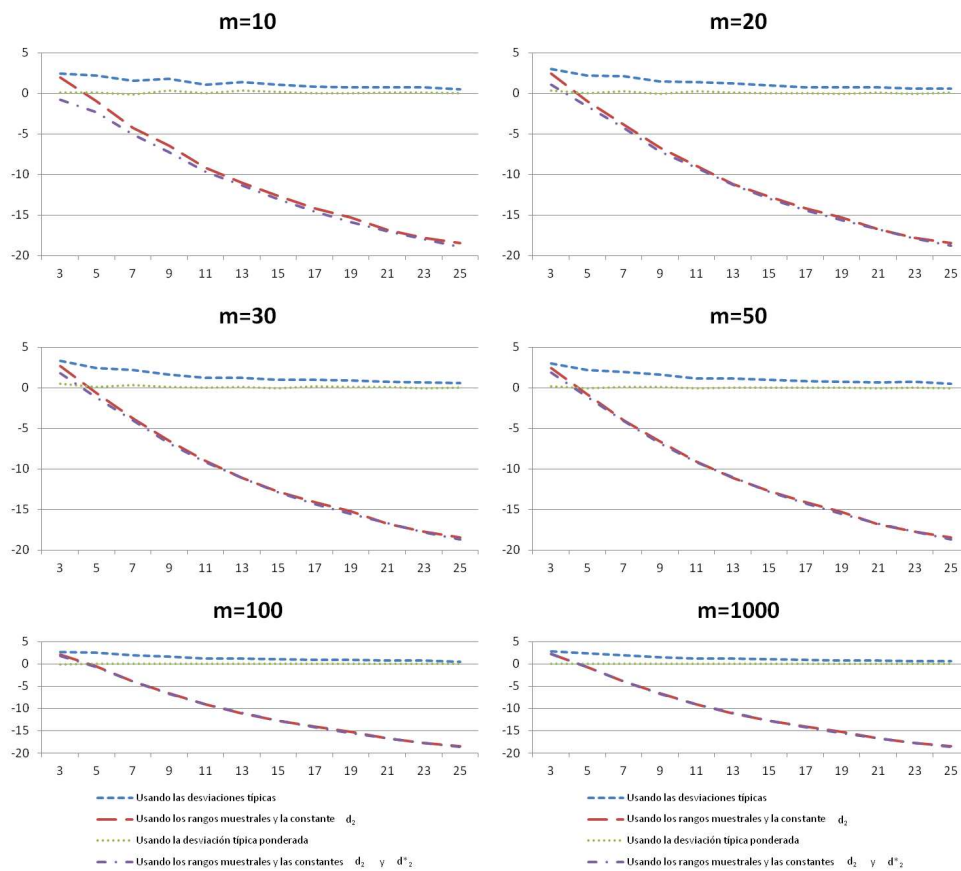


Figura 2.33: Valores de  $RB$  (%) de los estimadores  $\hat{\sigma}_R$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{\sigma}_{R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{\sigma}_S$  (usando la desviación típica muestral) y  $\hat{\sigma}_P$  (usando la desviación típica ponderada) de  $\sigma = 1$ . Datos seleccionados de la distribución Uniforme.



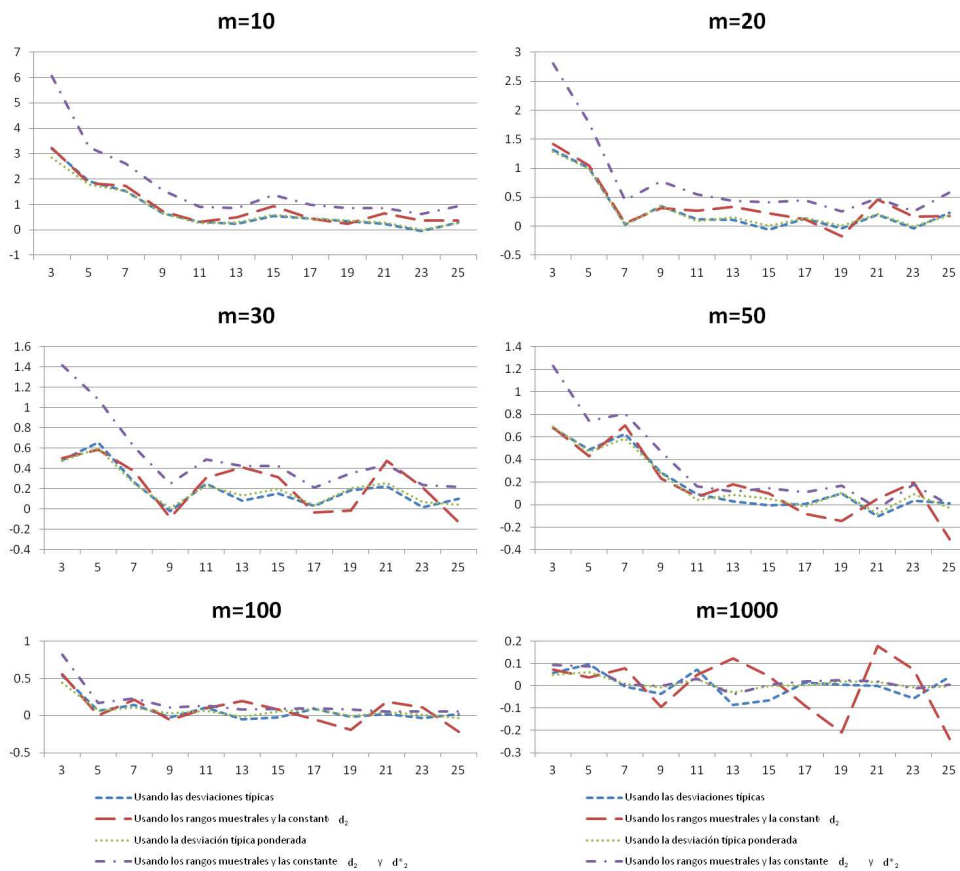


Figura 2.34: Valores de  $RB$  (%) de los estimadores  $\widehat{C}_{pk.R}$  (usando los rangos muestrales y la constante  $d_2$ ),  $\widehat{C}_{pk.R2}$  (usando los rangos muestrales y las constantes  $d_2$  and  $d_2^*$ ),  $\widehat{C}_{pk.S}$  (usando la desviación típica muestral) y  $\widehat{C}_{pk.P}$  (usando la desviación típica ponderada) de  $C_{pk} = 1$ . Datos seleccionados de la distribución Normal con  $\sigma = 1$  en un proceso no centrado.

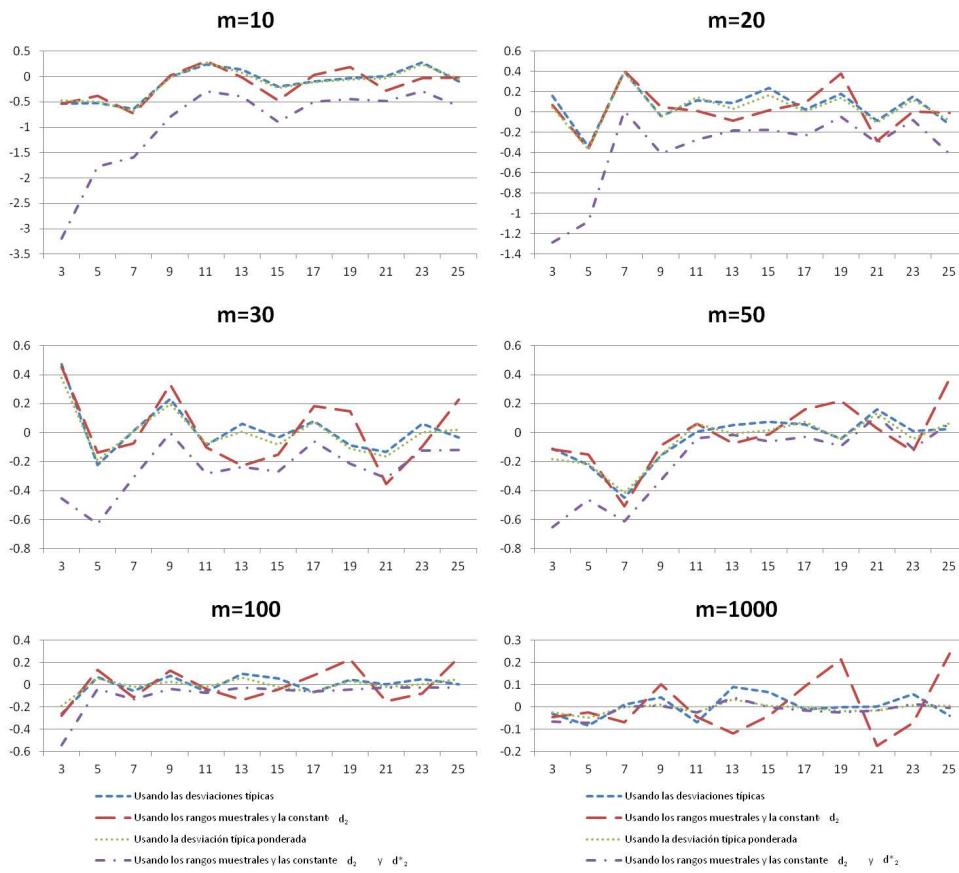


Figura 2.35: Valores de  $RB$  (%) de los estimadores  $\hat{\sigma}_R$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{\sigma}_{R2}$  (usando los rangos muestrales y las constantes  $d_2$  and  $d_2^*$ ),  $\hat{\sigma}_S$  (usando la desviación típica muestral) y  $\hat{\sigma}_P$  (usando la desviación típica ponderada) de  $\sigma = 1$ . Datos seleccionados de la distribución Normal en un proceso no centrado.

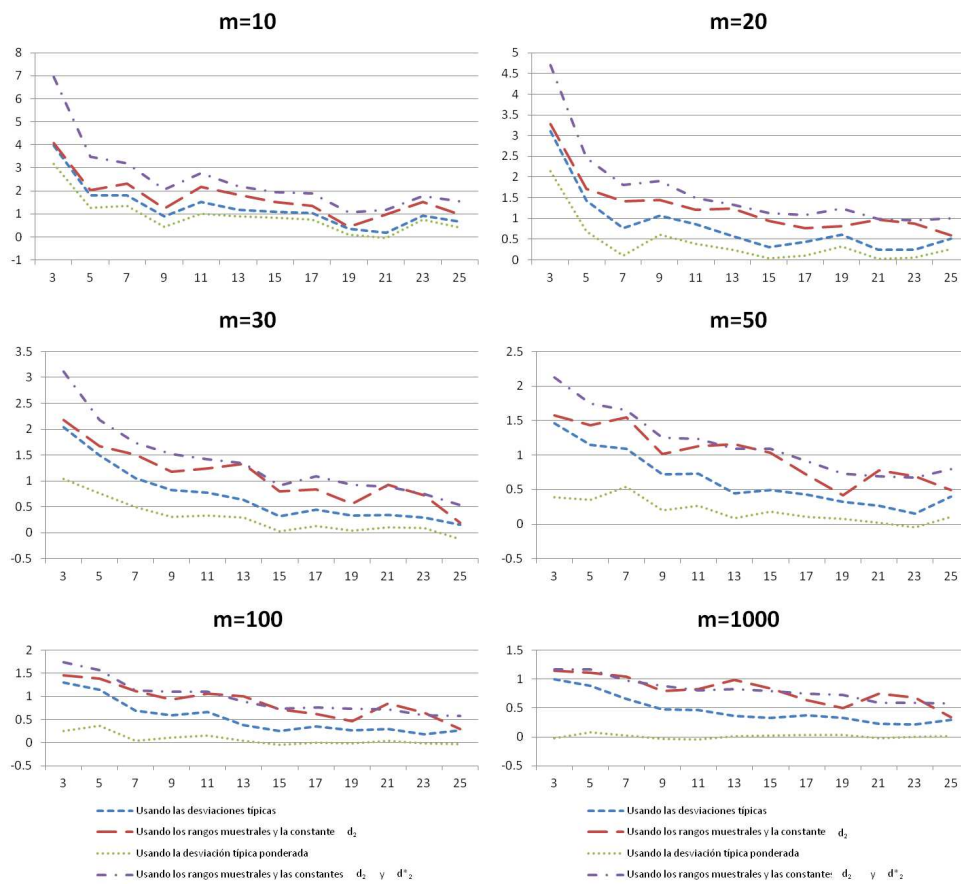


Figura 2.36: Valores de  $RB$  (%) de los estimadores  $\hat{C}_{p.R}$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{C}_{p.R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{C}_{p.S}$  (usando la desviación típica muestral) y  $\hat{C}_{p.P}$  (usando la desviación típica ponderada) de  $C_p = 1$ . Datos seleccionados de la distribución Gamma con  $\sigma = 3$ .

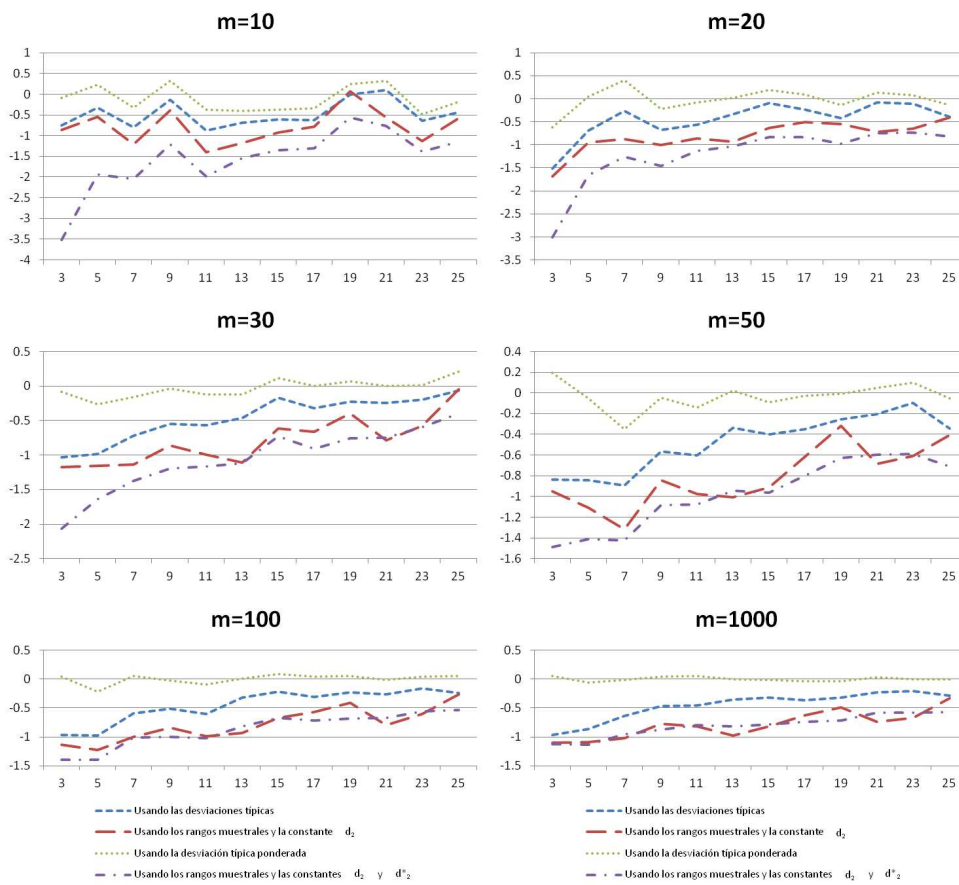


Figura 2.37: Valores de  $RB$  (%) de los estimadores  $\hat{\sigma}_R$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{\sigma}_{R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{\sigma}_S$  (usando la desviación típica muestral) y  $\hat{\sigma}_P$  (usando la desviación típica ponderada) de  $\sigma = 1$ . Datos seleccionados de la distribución Gamma.

trado y los datos proceden de la distribución Normal (Figuras 2.34 y 2.35). Tales resultados deben compararse con los resultados derivados de las Figuras 2.28 y 2.29 (datos seleccionados de la distribución Normal y el proceso es centrado). Se observa que los valores de  $RB$  basados en procesos no centrados (Figuras 2.34 y 2.35) son similares, respectivamente, a los valores de  $RB$  de las Figuras 2.28 y 2.29. Esta cuestión indica que el impacto en los diferentes estimadores de  $C_{pk}$  y  $\sigma$  no es relevante si se consideran procesos no centrados.

En las Figuras 2.38–2.45 se puede analizar la eficiencia, en términos de  $RRMSE$ , de los diferentes estimadores de  $C_p$  y  $\sigma$ . Recordamos que un estimador A es más eficiente empíricamente que un estimador B si el  $RRMSE$  del estimador A es menor que el  $RRMSE$  del estimador B. Destacamos que en ningún caso se habla de estimadores eficientes, puesto que el problema de obtener la cota de Frechét-Cramer-Rao en estimadores complejos tal como los descritos para la estimación de  $C_p$  tiene difícil solución. Como era de esperar, los diferentes estimadores son generalmente más eficientes cuando se incrementan los valores de  $n$  y  $m$ . La ganancia en la eficiencia aumenta cuando los valores de  $n$  están aumentando y  $n$  es pequeño, es decir, por lo general, se observa que el impacto del aumento del valor  $n$  es más pequeño cuando  $n$  es mayor que 11, ya que la pendiente de las distintas curvas es menor en esta situación. También se observó que los estimadores basados en los rangos de la muestra son menos precisos que sus competidores cuando  $n$  toma valores elevados. También se obtiene una ganancia importante en la eficiencia de los distintos estimadores cuando se incrementa el valor de  $m$ . En el caso de la distribución Uniforme (Figuras 2.42 y 2.43), la eficiencia de los estimadores basados en los rangos muestrales disminuye a medida que los valores de  $n$  aumentan.

Por último, destacamos que también se llevaron a cabo estudios de simulación Monte Carlo mediante el uso de las diferentes combinaciones derivadas de los valores  $\sigma = \{1, 3\}$  y  $C_p = \{1, 1,5\}$ , y se obtuvieron conclusiones similares en comparación con los resultados que se han descrito dentro de esta sección.

### 2.3.3. Conclusiones

El objetivo principal de este capítulo es analizar el comportamiento empírico de diferentes estimadores del índice de capacidad del proceso ( $C_p$ ) y asumiendo diferentes escenarios que pueden presentarse en la práctica. Para este propósito, se han llevado a cabo los estudios de simulación Monte Carlo, basados en diversos valores de: (i) la desviación típica del proceso; (ii) especifica-

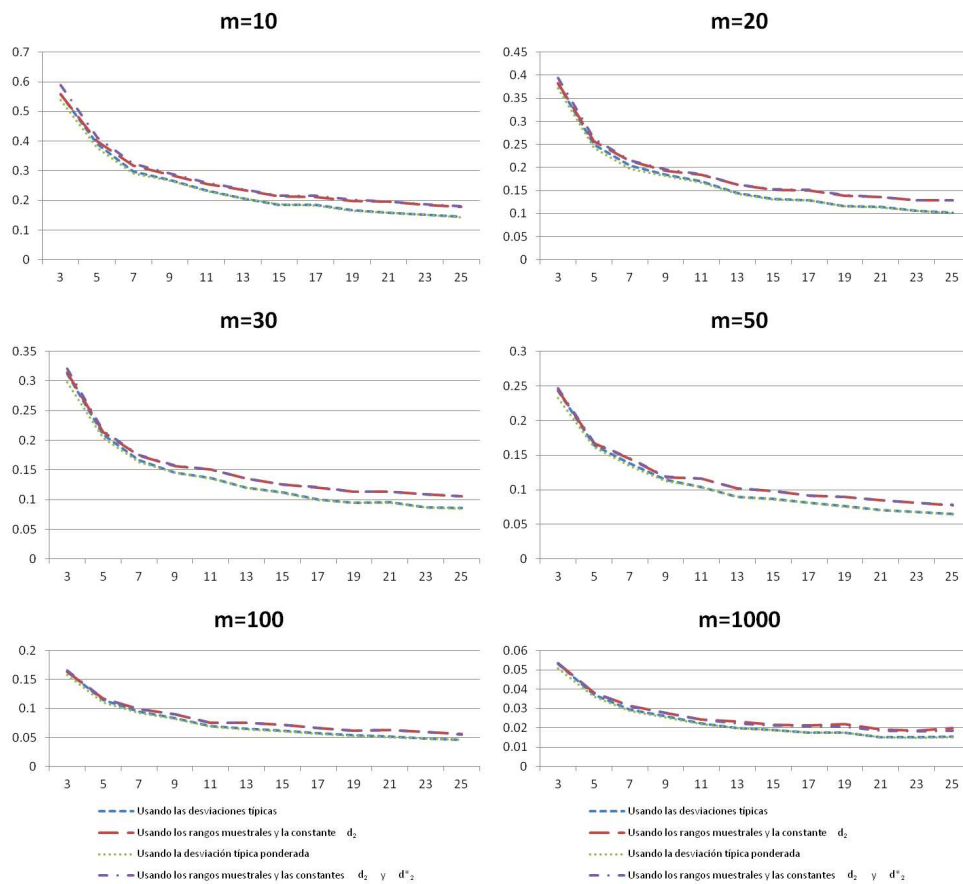


Figura 2.38: Valores de  $RRMSE$  (%) de los estimadores  $\hat{C}_{p.R}$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{C}_{p.R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{C}_{p.S}$  (usando la desviación típica muestral) y  $\hat{C}_{p.P}$  (usando la desviación típica ponderada) de  $C_p = 1$ . Datos seleccionados de la distribución Normal con  $\sigma = 1$ .

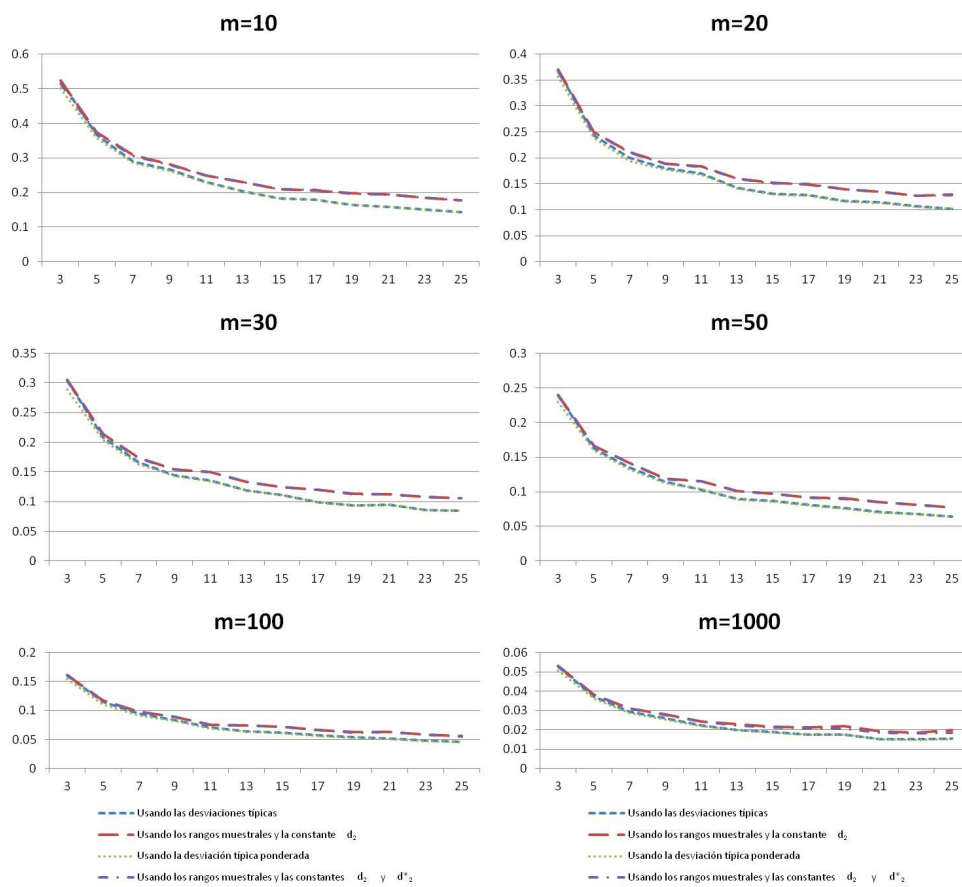


Figura 2.39: Valores de  $RRMSE$  (%) de los estimadores  $\hat{\sigma}_R$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{\sigma}_{R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{\sigma}_S$  (usando la desviación típica muestral) y  $\hat{\sigma}_P$  (usando la desviación típica ponderada) de  $\sigma = 1$ . Datos seleccionados de la distribución Normal.

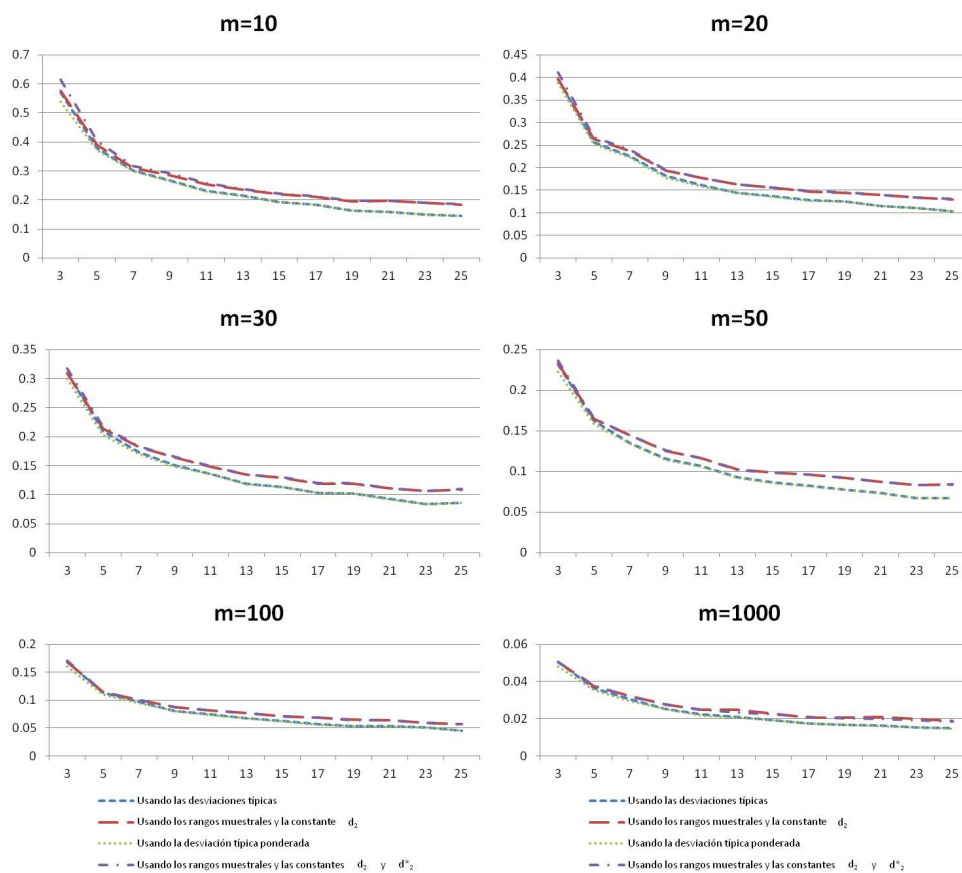


Figura 2.40: Valores de  $RRMSE$  (%) de los estimadores  $\hat{C}_{p,R}$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{C}_{p,R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{C}_{p,S}$  (usando la desviación típica muestral) y  $\hat{C}_{p,P}$  (usando la desviación típica ponderada) de  $C_p = 1$ . Datos seleccionados de la distribución Gamma con  $\sigma = 1$ .



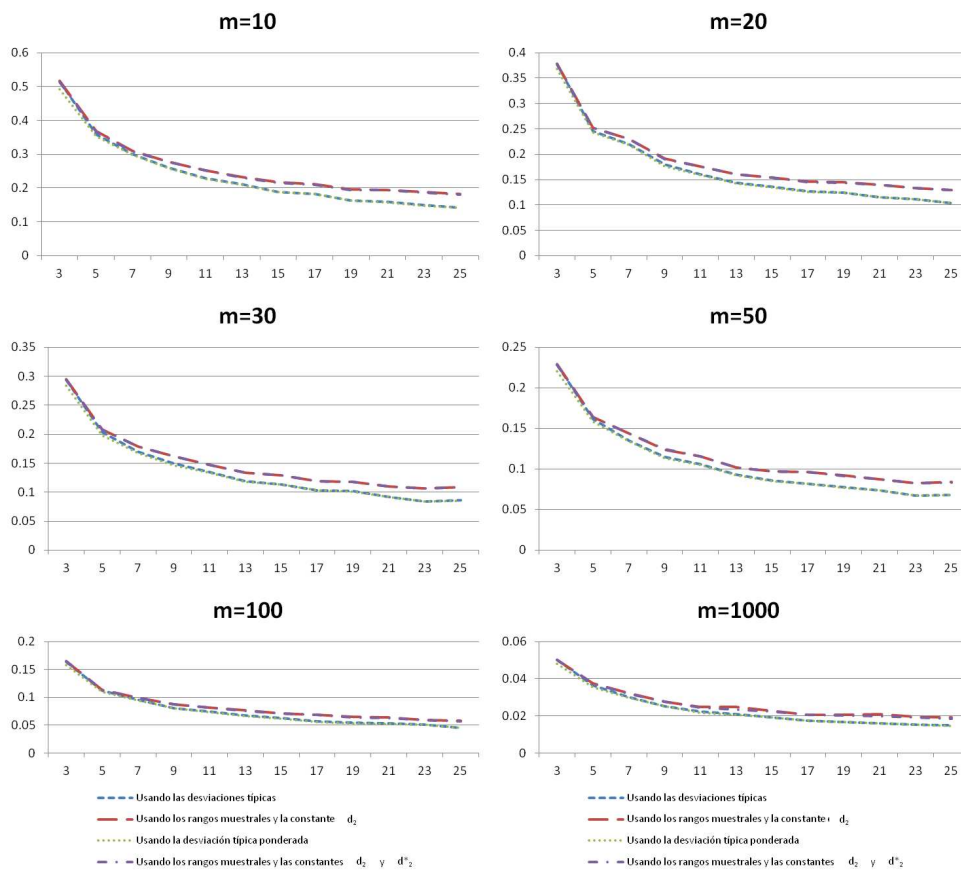


Figura 2.41: Valores de  $RRMSE$  (%) de los estimadores  $\hat{\sigma}_R$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{\sigma}_{R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{\sigma}_S$  (usando la desviación típica muestral) y  $\hat{\sigma}_P$  (usando la desviación típica ponderada) de  $\sigma = 1$ . Datos seleccionados de la distribución Gamma.

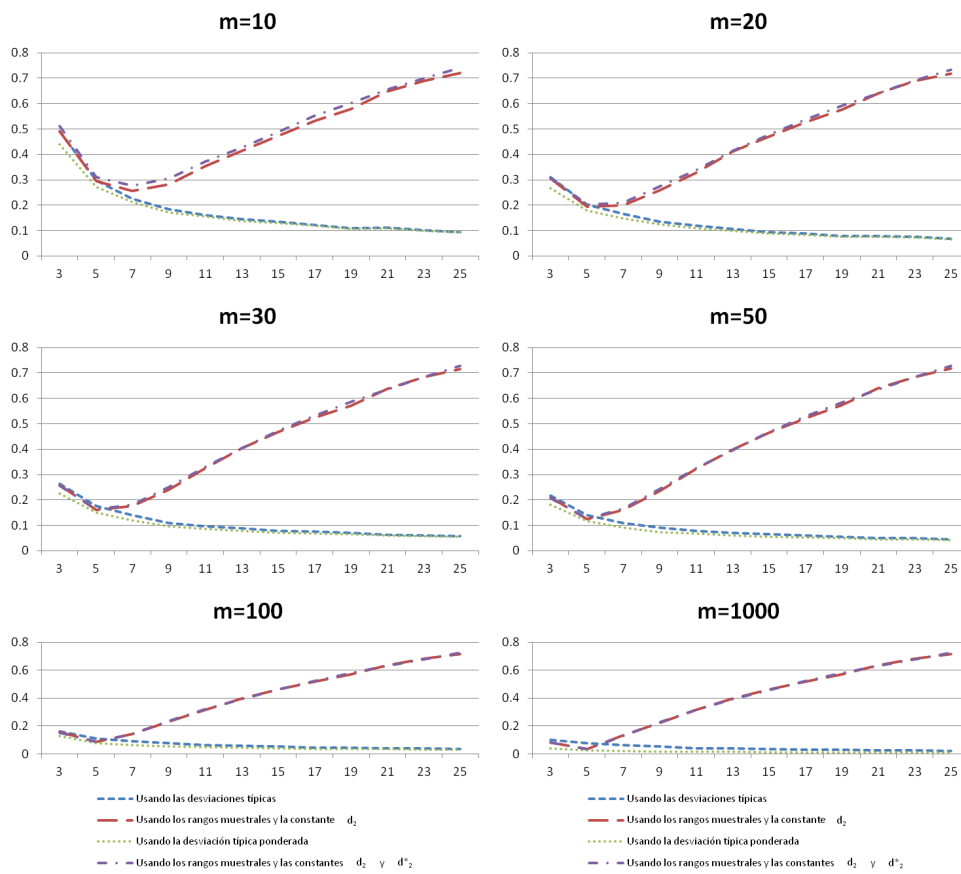


Figura 2.42: Valores de  $RRMSE$  (%) de los estimadores  $\hat{C}_{p,R}$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{C}_{p,R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{C}_{p,S}$  (usando la desviación típica muestral) y  $\hat{C}_{p,P}$  (usando la desviación típica ponderada) de  $C_p = 1$ . Datos seleccionados de la distribución Uniforme con  $\sigma = 1$ .

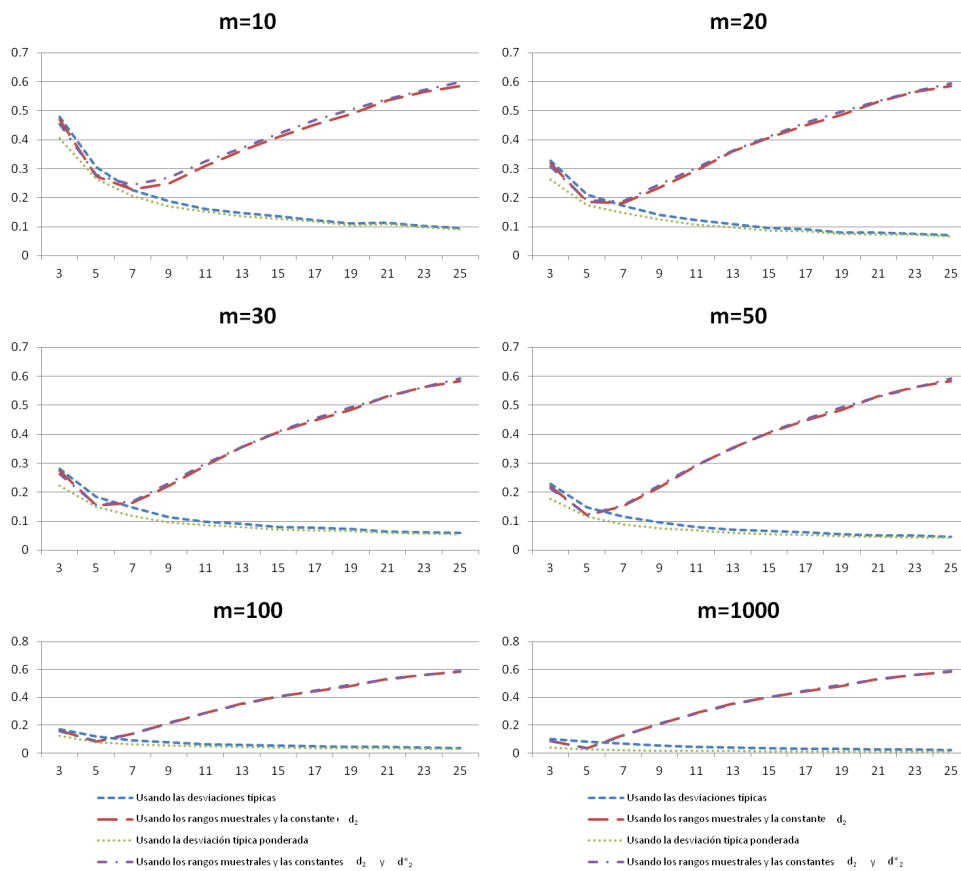


Figura 2.43: Valores de  $RRMSE$  (%) de los estimadores  $\hat{\sigma}_R$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{\sigma}_{R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{\sigma}_S$  (usando la desviación típica muestral) y  $\hat{\sigma}_P$  (usando la desviación típica ponderada) de  $\sigma = 1$ . Datos seleccionados de la distribución Uniforme.

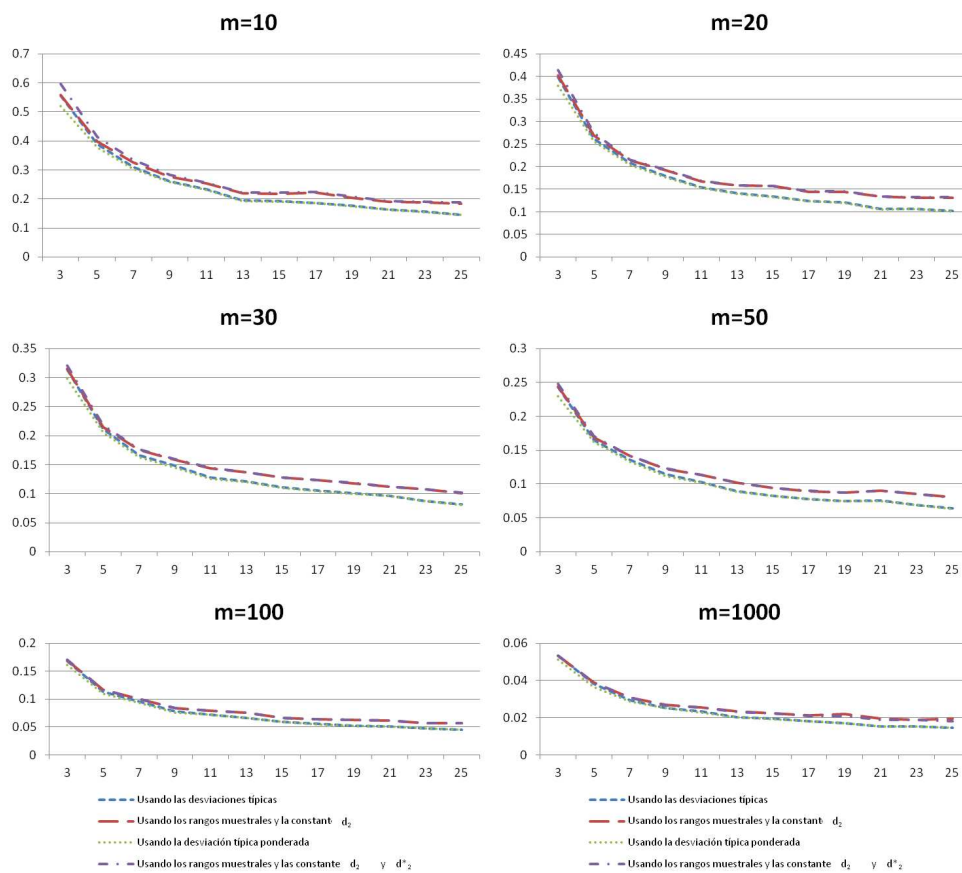


Figura 2.44: Valores de  $RRMSE$  (%) de los estimadores  $\hat{C}_{pk.R}$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{C}_{pk.R2}$  (usando los rangos muestrales y las constantes  $d_2$  and  $d_2^*$ ),  $\hat{C}_{pk.S}$  (usando la desviación típica muestral) y  $\hat{C}_{pk.P}$  (usando la desviación típica ponderada) de  $C_{pk} = 1$ . Datos seleccionados de la distribución Normal con  $\sigma = 1$  en un proceso no centrado.

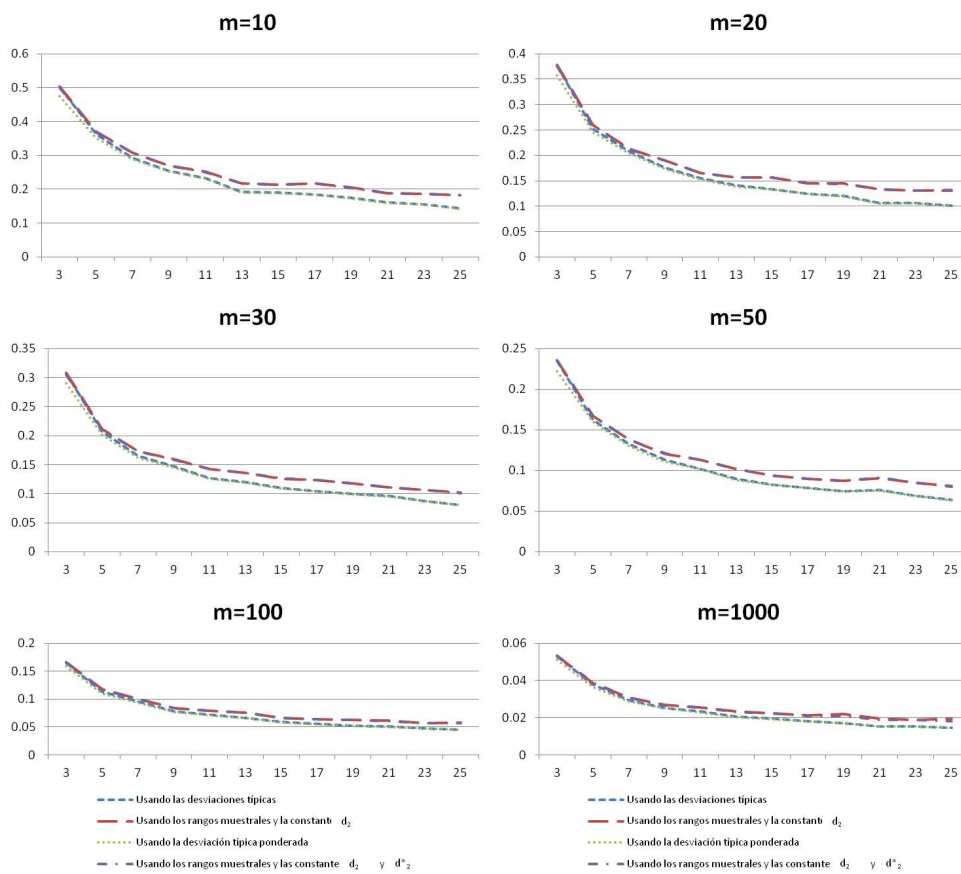


Figura 2.45: Valores de  $RRMSE$  (%) de los estimadores  $\hat{\sigma}_R$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{\sigma}_{R2}$  (usando los rangos muestrales y las constantes  $d_2$  and  $d_2^*$ ),  $\hat{\sigma}_S$  (usando la desviación típica muestral) y  $\hat{\sigma}_P$  (usando la desviación típica ponderada) de  $\sigma = 1$ . Datos seleccionados de la distribución Normal en un proceso no centrado.

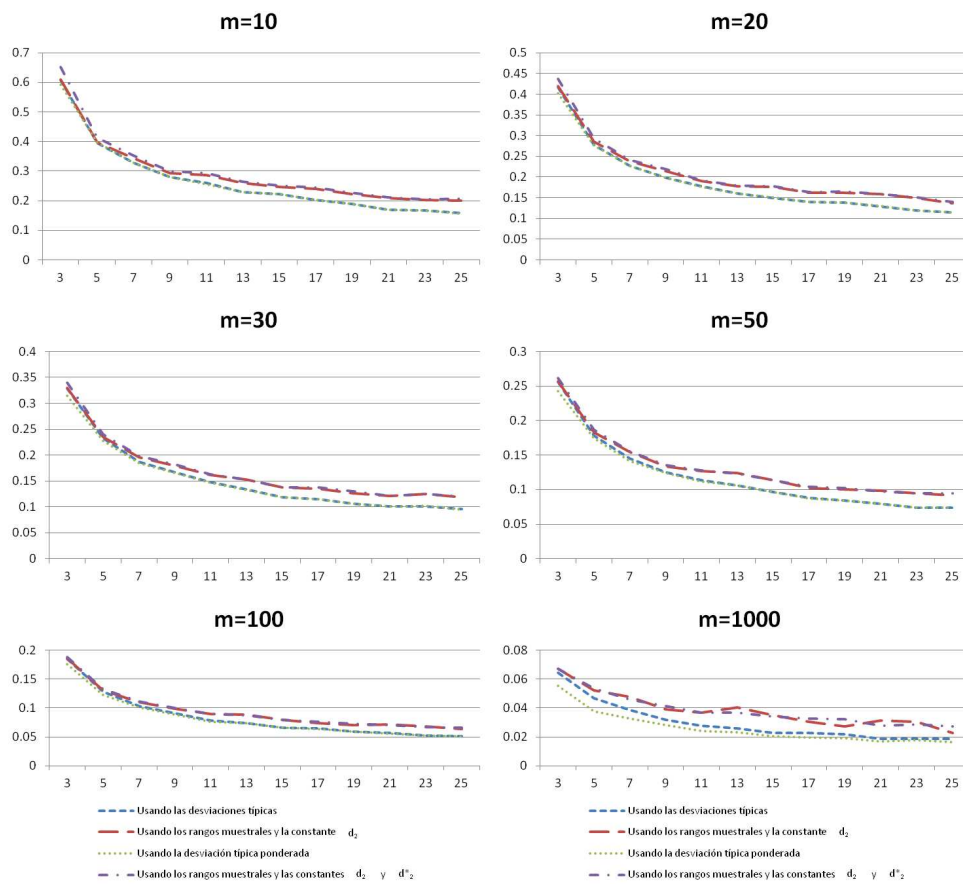


Figura 2.46: Valores de  $RRMSE$  (%) de los estimadores  $\hat{C}_{p,R}$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{C}_{p,R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{C}_{p,S}$  (usando la desviación típica muestral) y  $\hat{C}_{p,P}$  (usando la desviación típica ponderada) de  $C_p = 1$ . Datos seleccionados de la distribución Gamma con  $\sigma = 3$ .

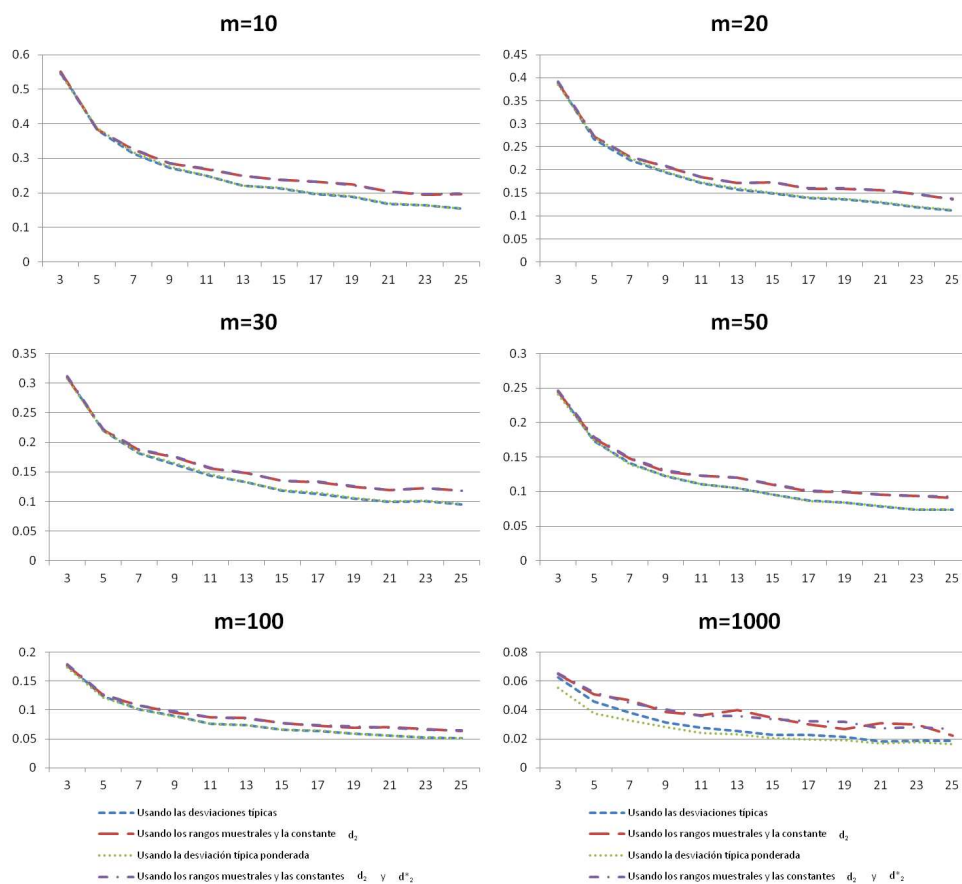


Figura 2.47: Valores de  $RRMSE$  (%) de los estimadores  $\hat{\sigma}_R$  (usando los rangos muestrales y la constante  $d_2$ ),  $\hat{\sigma}_{R2}$  (usando los rangos muestrales y las constantes  $d_2$  y  $d_2^*$ ),  $\hat{\sigma}_S$  (usando la desviación típica muestral) y  $\hat{\sigma}_P$  (usando la desviación típica ponderada) de  $\sigma = 3$ . Datos seleccionados de la distribución Gamma.

ciones, o, de manera similar, los valores del verdadero índice de capacidad del proceso; (iii) tamaño de muestras  $n$  y; (iv) número de muestras  $m$  utilizadas para obtener los diferentes estimadores. También se consideraron diferentes distribuciones probabilísticas para analizar el impacto del incumplimiento del supuesto de normalidad en los diferentes estimadores de  $C_p$  y  $\sigma$ . Por último, se han considerado procesos no centrados, y se ha analizado el comportamiento empírico de los diferentes estimadores en esta situación. Los resultados se comparan en términos de sesgo y eficiencia, medidos a través de las medidas empíricas  $RB$  y  $RRMSE$ , respectivamente.

En primer lugar, se observan grandes sesgos cuando  $n$  es menor a 5. Los sesgos de los diferentes estimadores no sufren un impacto significativo cuando  $n$  es mayor que 5. Sin embargo, la variabilidad de los sesgos del estimador basado en los rangos de la muestra es mayor en comparación con los estimadores alternativos. Como era de esperar, los diferentes estimadores son más eficientes cuando ambos valores de  $n$  y  $m$  incrementan. Las figuras obtenidas en este estudio se podrían utilizar para analizar el impacto en los diferentes estimadores a medida que aumentamos los dos valores de  $n$  y  $m$ . Para valores elevados de  $n$ , los estimadores basados en los rangos muestrales son menos precisos que sus competidores. Este problema puede deberse al hecho de que los sesgos tienen una gran variabilidad en esta situación. También se analizó el comportamiento empírico de los diferentes estimadores cuando los datos proceden de las distribuciones Gamma y Uniforme. Se observan resultados similares cuando se considera la distribución Gamma. Sin embargo, también se observó que la distribución Uniforme tiene un impacto importante sobre el comportamiento de los diferentes estimadores basados en los rangos muestrales.

En resumen, los resultados derivados de los estudios de simulación Monte Carlo indican que los estimadores basados en los rangos muestrales son ligeramente menos precisos que sus competidores, especialmente cuando el tamaño de la muestra,  $n$ , aumenta. Tales estimadores pueden tener un mal comportamiento cuando no se satisface el supuesto de normalidad. En particular, los estimadores basados en los rangos muestrales tienen un comportamiento muy pobre cuando se utilizan datos generados a partir de la distribución Uniforme y el tamaño muestral  $n$  es elevado. Los diversos estimadores pueden tener importantes sesgos cuando los tamaños de las muestras son menores que 5.





# Capítulo 3

## Aportaciones relacionadas con la función $c_4[\cdot]$ .

### 3.1. Introducción

La estimación puntual de parámetros es una técnica ampliamente utilizada, y también tiene un interés muy destacado en economía y empresa. Su objetivo principal es inferir sobre un parámetro poblacional desconocido. Asimismo, los estimadores puntuales son necesarios para otras técnicas de inferencia como la construcción de intervalos de confianza y contrastes de hipótesis (Herrerías et al. 2012 pp.77 y 107). Por ejemplo, la encuesta de población activa (EPA) llevada a cabo por el Instituto Nacional de Estadística (INE) tiene el objetivo principal de realizar estimaciones de la población española en relación con el mercado de trabajo: activos, ocupados, parados e inactivos. Otro ejemplo en el ámbito de las Ciencias Sociales es la Encuesta Europea sobre ingresos y condiciones de vida (European Union Statistics on Income and Living Condition, EU-SILC). Esta encuesta se realiza a nivel de la Unión Europea por la Oficina Europea de Estadística (EUROSTAT), y uno de sus numerosos objetivos es estimar la proporción de pobres de cada país de la Unión Europea. Los estimadores puntuales también son fundamentales en disciplinas como la econometría. Por ejemplo, los coeficientes de regresión asociados a un modelo econométrico se consideran constantes desconocidas que deben ser estimadas y, para ello, se utilizan los datos procedentes de una muestra (Gujarati 2009, Wooldridge 2015).

Algunos parámetros tienen una notable importancia en determinadas dis-

ciplinas. En concreto, la desviación típica tiene especial relevancia en el SQC, tal como se ha podido comprobar en capítulos anteriores. En efecto, en dichos capítulos se comentó que la variabilidad de los procesos productivos se asocia tradicionalmente a la desviación típica de la característica de calidad, y se recurre a una estimación puntual de este parámetro cuando éste es desconocido. Destacamos que esta es la situación más común en la práctica. La estimación puntual de la desviación típica tiene numerosas aplicaciones, tal como la obtención de ciertos límites: límites de variabilidad natural del proceso; límites de control; etc. Además, la estimación puntual de la desviación típica es requerida para obtener estimadores de los índices de capacidad (objeto de estudio de esta investigación), gráficos de control, etc.

En resumen, los parámetros presentes en una multitud de disciplinas y estudios suelen ser desconocidos pero estimados a partir de datos procedentes de muestras aleatorias, y este hecho pone de manifiesto la importancia del uso de estimadores puntuales. Tales estimadores deben gozar de una serie de propiedades deseables. En este sentido, una propiedad deseable de los estimadores puntuales es la insesgadez. Por ejemplo, es bien conocido que estadísticos como la media muestral o la cuasivarianza muestral son estimadores insesgados de sus respectivos parámetros, es decir, la media poblacional y la varianza poblacional, respectivamente.

Siguiendo la notación usada en capítulos anteriores, denotaremos por  $X$  a la característica de calidad de un determinado artículo en un proceso de producción, y asumiremos que dicha característica de calidad sigue una distribución Normal con media  $\mu$  y desviación típica  $\sigma$ , es decir,  $X \rightsquigarrow N(\mu, \sigma)$ . Se denota como  $x_1, \dots, x_n$  a cada una de las observaciones de la característica de calidad  $X$  pertenecientes a una muestra de tamaño  $n$ . Los estimadores de la media poblacional ( $\mu$ ) y la varianza poblacional ( $\sigma^2$ ) se definen como

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

y

$$\hat{\sigma}^2 = S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2,$$

respectivamente. Como se argumentaba con anterioridad, los valores esperados de ambos estimadores son

$$E[\bar{x}] = \mu$$

y

$$E[\widehat{\sigma}^2] = \sigma^2,$$

respectivamente. A pesar de que la cuasivarianza muestral  $\widehat{\sigma}^2$  es un estimador insesgado de la varianza poblacional, la cuasideviación típica es un estimador sesgado de la desviación típica poblacional, es decir,

$$E[\widehat{\sigma}] \neq \sigma,$$

donde

$$\widehat{\sigma} = S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2}.$$

Esta pérdida de la propiedad de insesgadez se debe a que la raíz cuadrada no es una función o transformación lineal, lo cual sí conservaría dicha propiedad de insesgadez. No obstante, el valor esperado de  $\widehat{\sigma}$  puede obtenerse mediante la conocida distribución dada por

$$\frac{(n-1)\widehat{\sigma}^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2,$$

tal como puede comprobarse en el Resultado 3.1 que se presenta a continuación. A partir de este resultado es posible obtener fácilmente un estimador insesgado de la desviación típica poblacional. Este estimador insesgado se suele expresar en la literatura a través de la función denominada  $c_4[\cdot]$ , y la cual depende del tamaño muestral. En capítulos anteriores se ha hecho uso de esta función, pero en el presente capítulo se define de manera formal esta función, se razona su expresión analítica, y se obtienen nuevas expresiones para esta función que pueden resultar útiles en la práctica. Recordamos que los valores de  $c_4[\cdot]$  para determinados tamaños muestrales están, en general, disponibles en los textos sobre SQC, o bien pueden calcularse de forma simple para tamaños muestrales pequeños en diversos programas estadísticos. Sin embargo, para tamaños muestrales elevados su cálculo no resulta posible a partir de su expresión original, puesto que depende del cociente de dos funciones  $\Gamma(\cdot)$  que no se pueden simplificar a partir del factorial de números enteros (véase la Sección 3.2). La función  $\Gamma(\cdot)$  se define como

$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} dx \quad \text{si } \forall \alpha > 0.$$

Además, determinados programas estadísticos, como por ejemplo *R*, no permiten el cálculo de  $c_4[\cdot]$  en esta situación de tamaños elevados.

La idea principal de este capítulo es proponer nuevas expresiones analíticas para determinar el valor de  $c_4[\cdot]$  que no dependan de las funciones  $\Gamma(\cdot)$  anteriormente comentadas. Por los motivos ya expuestos, parece conveniente la disponibilidad de expresiones para  $c_4[\cdot]$  que puedan obtenerse para cualquier tamaño muestral, dando el valor exacto, sin hacer uso de aproximaciones, y sin necesidad de evaluar previamente la disponibilidad del valor de  $c_4[\cdot]$ . En otras palabras, tales expresiones permitirán programar en programas estadísticos funciones basadas en  $c_4[\cdot]$  sin restricciones para el tamaño de muestra considerado en el correspondiente estudio. En particular, se han propuesto dos nuevas expresiones para el cálculo de  $c_4[\cdot]$ . La primera expresión hace posible el cálculo de  $c_4[\cdot]$  para cualquier tamaño de muestra, pero relacionándolo con el valor de la misma función  $c_4[\cdot]$  evaluada en otro número entero inferior al tamaño de muestra. La segunda expresión resulta más práctica, puesto que hace posible el cálculo de  $c_4[\cdot]$  sin necesidad de usar el valor de  $c_4[\cdot]$  para otro número entero. Por último, destacamos que ambas expresiones calculan el valor de  $c_4[\cdot]$  como producto de múltiples fracciones, y se evita la presencia de valores infinitos que provoquen una indeterminación en el cálculo de  $c_4[\cdot]$  en el caso de valores muestrales elevados, tal como ocurre en la expresión que define  $c_4[\cdot]$ .

Por tanto, en este capítulo se aborda el objetivo O11 descrito en la Sección 1.3, y para ello, se estructura este capítulo de la siguiente forma. En la Sección 3.2 se define y se razona apropiadamente la función  $c_4[\cdot]$ . En la Sección 3.3 se presentan las nuevas expresiones desarrolladas para esta función básica en el contexto del SQC. En la Sección 3.4 se exponen las principales conclusiones derivadas de este capítulo.

## 3.2. Definición y justificación de la función $c_4[\cdot]$

En primer lugar, se razona en el Resultado 3.1 la definición de la función  $c_4[\cdot]$ . Como puede observarse, esta función puede justificarse como un coeficiente corrector que permite obtener un estimador insesgado de la desviación típica poblacional.

**Resultado 3.1.** El estimador

$$\hat{\sigma}_{c_4} = \frac{S}{c_4[n]}$$

de la desviación típica poblacional  $\sigma$  es insesgado, donde

$$c_4[n] = \frac{\sqrt{2}}{\sqrt{n-1}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}. \quad (3.1)$$

□

### Demostración

En el caso de variables aleatorias independientes e idénticamente distribuidas (i.i.d.) se conoce que el estadístico

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2}$$

se distribuye como una distribución chi-cuadrado, es decir,

$$\frac{(n-1)\hat{\sigma}^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2 \equiv \gamma\left(\frac{n-1}{2}, 2\right), \quad (3.2)$$

y por tanto, su esperanza matemática viene dada por

$$E\left[\frac{(n-1)\hat{\sigma}^2}{\sigma^2}\right] = n-1.$$

Dado que el interés es obtener el valor esperado del estadístico  $\sqrt{n-1}\hat{\sigma}/\sigma$ , por simplicidad definimos la variable  $Y$  como

$$Y = \frac{(n-1)\hat{\sigma}^2}{\sigma^2},$$

por lo que el siguiente paso es la obtención del valor esperado de la raíz cuadrada de  $Y$ , es decir:

$$E[Y^{1/2}] = \int_{-\infty}^{+\infty} y^{1/2} f(y) dy.$$

Haciendo uso de la función de densidad de la distribución chi-cuadrado nos quedaría:

$$\begin{aligned} E[Y^{1/2}] &= \int_0^{\infty} y^{1/2} \frac{1}{2^{(n-1)/2} \Gamma((n-1)/2)} y^{((n-1)/2)-1} e^{-y/2} dy = \\ &= \frac{1}{2^{(n-1)/2} \Gamma((n-1)/2)} \int_0^{\infty} y^{(n/2)-1} e^{-y/2} dy. \end{aligned}$$

La integral anterior se puede resolver mediante la técnica popularmente conocida como “cambio de variable”, y donde se define

$$t = \frac{y}{2},$$

por lo que

$$dt = \frac{1}{2}dy.$$

Aplicando en la última integral este cambio nos queda:

$$\begin{aligned} E[Y^{1/2}] &= \frac{1}{2^{(n-1)/2}\Gamma((n-1)/2)} \int_0^\infty (2t)^{(n/2)-1} e^{-t} 2dt = \\ &= \frac{2 \cdot 2^{(n/2)-1}}{2^{(n-1)/2}\Gamma((n-1)/2)} \int_0^\infty (t)^{(n/2)-1} e^{-t} dt = \\ &= \sqrt{2} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)}. \end{aligned}$$

En otras palabras,

$$E[Y^{1/2}] = E\left[\frac{\sqrt{n-1}\hat{\sigma}}{\sigma}\right] = \sqrt{2} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)},$$

lo cual nos permite conocer el valor esperado del estimador  $\hat{\sigma} = S$ :

$$E[S] = \frac{\sqrt{2}}{\sqrt{n-1}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \sigma.$$

Definiendo  $c_4[n]$  mediante la expresión (3.1), el valor esperado de la cuasidesviación típica muestral viene dado por  $E[S] = c_4[n]\sigma$ , por lo que un estimador insesgado de la desviación típica poblacional  $\sigma$  puede definirse fácilmente como

$$\hat{\sigma}_{c4} = \frac{S}{c_4[n]},$$

puesto que

$$E[\hat{\sigma}_{c4}] = E\left[\frac{S}{c_4[n]}\right] = \frac{1}{c_4[n]} E[S] = \frac{1}{c_4[n]} c_4[n]\sigma = \sigma.$$

□

Una vez conocido que la cuasidesviación típica muestral  $S$  es un estimador sesgado de la desviación típica poblacional, el siguiente paso es determinar el valor teórico de este sesgo, lo cual se analiza en el siguiente resultado.

**Resultado 3.2.** El sesgo del estimador  $\hat{\sigma} = S$  viene dado por

$$B[S] = \left( \frac{\sqrt{2}}{\sqrt{n-1}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)} - 1 \right) \sigma$$

□

### Demostración

A partir del Resultado 3.1 puede comprobarse fácilmente que el sesgo de  $\hat{\sigma} = S$  viene dado por

$$B[S] = E[S] - \sigma = \frac{\sqrt{2}}{\sqrt{n-1}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)} \sigma - \sigma = \left( \frac{\sqrt{2}}{\sqrt{n-1}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)} - 1 \right) \sigma.$$

□

Se puede comprobar que el término que acompaña a la desviación típica muestral en la expresión del sesgo de  $S$  es negativo, es decir,

$$\frac{\sqrt{2}}{\sqrt{n-1}} \frac{\Gamma(n/2)}{\Gamma((n-1)/2)} - 1 < 0,$$

lo que implica que el sesgo del estimador  $S$  es negativo. En la Figura 3.1 puede observarse el sesgo de la cuasidesviación típica muestral para distintos tamaños muestrales, y para un proceso de producción con desviación típica  $\sigma = 1$ .

Destacamos que la función  $c_4[\cdot]$  se aproxima a 1 a medida que aumenta el tamaño muestral. Esto implica que el factor de corrección del sesgo aplicado sobre el estimador  $S$ ,  $1/c_4[n]$ , es menor a medida que se incrementa el tamaño de muestra. Dicho de otro modo, la cuasidesviación típica muestral es un estimador asintóticamente insesgado, por lo que a medida que aumenta el tamaño muestral, este estimador requiere una corrección menor. Este hecho puede también observarse en los resultados mostrados en la Tabla 2.2 del Capítulo 2. En dicha tabla, si se comparan los distintos estimadores de la desviación típica basados en poblaciones infinitas, se aprecia, por un lado, que el estimador sesgado de la desviación típica presenta sesgos negativos, aunque mejora su comportamiento en términos de sesgo relativo a medida que aumenta el tamaño muestral. Por otro lado, los sesgos relativos del estimador insesgado están próximos a 0, lo cual refleja la importancia y el efecto que puede tener el uso de la función  $c_4[\cdot]$  en el problema de la estimación de la desviación típica poblacional. En definitiva, a partir de los resultados teóricos y empíricos parece razonable hacer uso del estimador insesgado de la desviación típica poblacional, especialmente en el caso de muestras con tamaño pequeño, lo cual puede añadir mayor precisión al análisis realizado.

En determinados programas estadísticos, la expresión (3.1) que define la función  $c_4[\cdot]$  no puede obtenerse para tamaños muestrales elevados, puesto que depende del cálculo de la función  $\Gamma(\cdot)$  en dos valores diferentes, y es conocido



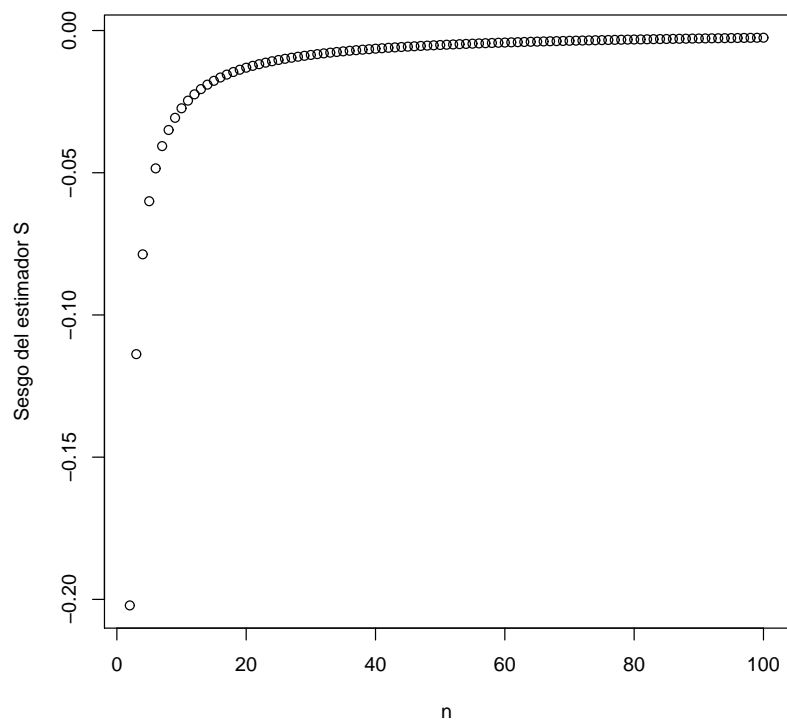


Figura 3.1: Sesgo teórico de la cuasidesviación típica muestral en un proceso con desviación típica  $\sigma = 1$  y para distintos tamaños de muestra.

que dicha función tiende a infinito a medida que aumenta el tamaño de la muestra. Una opción para su cálculo podría ser simplificar después de transformar las dos funciones  $\Gamma(\cdot)$  en números factoriales. Sin embargo, según la definición dada en la ecuación (3.1), puede observarse que no resulta posible en esta expresión la presencia simultánea de números enteros en el numerador y en el denominador. En consecuencia, a partir de un determinado tamaño de muestra, el valor asignado a la función  $\Gamma(\cdot)$  en numerosos programas estadísticos es infinito, por lo que tales programas no asignan un valor a la función  $c_4[\cdot]$  en esta situación. A modo de ejemplo, el software R obtiene una solución numérica de la función  $\Gamma(\cdot)$  hasta el valor 171,6144, por lo que este programa estadístico no aporta el valor de  $c_4[\cdot]$  para tamaños muestrales superiores a 343. Por los motivos expuestos, puede resultar relevante la disponibilidad de expresiones alternativas para la función  $c_4[\cdot]$ , las cuales se plantean a continuación en la Sección 3.3.

### 3.2.1. Análisis de la distribución de la cuasivarianza muestral en poblaciones finitas e infinitas

Los Resultados 3.1 y 3.2 asumen el cumplimiento de la distribución del estadístico definido en la expresión (3.2). Por ejemplo, esta distribución será válida en el caso de muestras procedentes de poblaciones infinitas, pero no se cumple desde un punto de vista teórico en el caso de muestras procedentes de poblaciones finitas. Existen multitud de escenarios en el contexto del SQC en los que las muestras pueden ser seleccionadas de poblaciones finitas. Por ejemplo, el muestreo de aceptación de lotes son un conjunto de herramientas estadísticas dentro del SQC basadas en muestras extraídas de lotes finitos de productos. Destacamos que en este trabajo se realizan varios estudios comparativos entre poblaciones finitas e infinitas. Por los motivos expuestos, y a modo de ilustración, a continuación se presenta un estudio basado en gráficos QQ, y en el cual se compara de forma numérica la distribución empírica del estadístico  $Y$  con su correspondiente distribución teórica. Para cada caso en este estudio, la distribución empírica está basada en un total de 100 muestras seleccionadas de una población Normal. Las muestras se seleccionan tanto en el caso de poblaciones finitas como infinitas, de modo que se puedan observar las diferencias entre ambas situaciones y se puedan extraer conclusiones. Destacamos que este estudio se corresponde con el objetivo O6 descrito en la Sección 1.3.

La Figura 3.2 contiene los gráficos QQ cuando el estadístico  $Y$  se basa en muestras obtenidas de una población infinita, y por tanto se dan las

circunstancias para que dicho estadístico se distribuya según la distribución  $\chi_{n-1}^2$ . Los gráficos QQ de la Figura 3.2 están basados en muestras de tamaño  $n = \{50, 100, 150, 200, 300, 500\}$ . Adicionalmente, se ha analizado la distribución del estadístico cuando las muestras se seleccionan de poblaciones finitas, es decir, no se cumple el Supuesto A2. En este caso de poblaciones finitas se han diferenciado dos escenarios: cuando la fracción de muestreo ( $f = n/N$ ) es elevada; y cuando  $f$  está próxima a 0. Se consideran poblaciones de tamaño  $N = 500$  y muestras con tamaño  $n = \{50, 100, 150, 200, 250\}$  para el caso de fracciones de muestreo elevadas, y las cuales toman los valores  $f = \{0,01, 0,1, 0,2, 0,3, 0,4, 0,5\}$ . Para el caso de fracciones pequeñas se consideran poblaciones con tamaño  $N = 10000$ , y muestras con tamaño  $n = \{10, 100, 200, 300, 400, 500\}$ , lo que implica que las fracciones de muestreo vienen dadas por  $f = \{0,001, 0,01, 0,02, 0,03, 0,04, 0,05\}$ . Los resultados derivados de poblaciones finitas se muestran en las Figuras 3.3 y 3.4. Como era de esperar y según los gráficos QQ, la distribución empírica se aproxima en gran medida a la distribución teórica en el caso de poblaciones infinitas (Figura 3.2). Para poblaciones finitas con fracciones de muestreo elevadas (Figura 3.3), se observa que a medida que la fracción de muestreo aumenta, la distribución empírica del estadístico se distancia más de la teórica. En particular, la distancia puede ser considerable para fracciones de muestreo superiores al 20 %. Por último, se observa una buena aproximación entre las distribuciones teóricas y empíricas en el caso de fracciones de muestreo pequeñas, lo que avalaría un buen ajuste a la distribución  $\chi_{n-1}^2$ .

### 3.3. Nuevas expresiones para el cálculo de la función $c_4[\cdot]$

En esta sección se presentan nuevas expresiones para el cálculo del término  $c_4[\cdot]$  que satisfacen algunas propiedades deseables, tal como el hecho de ser una expresión directa e inmediata para tamaños de muestra elevados, evitando fórmulas recursivas. A continuación, en el Resultado 3.4 se puede comprobar que el valor de  $c_4[\cdot]$  para un tamaño muestral  $n + 1$  se puede obtener en función de  $c_4[\cdot]$  pero evaluado en  $n$ . Esta relación permitiría obtener el valor de  $c_4[\cdot]$  para cualquier tamaño muestral, pero se trataría de una expresión recursiva, que no permite obtener su valor de forma directa, y poco recomendada en programas estadísticos debido al mayor tiempo en ejecución que requiere. Por ejemplo, en el caso de usar el programa estadístico R, si se quiere obtener el valor de  $c_4[500]$ , debería calcularse previamente  $c_4[499]$ . Para obtener esta última

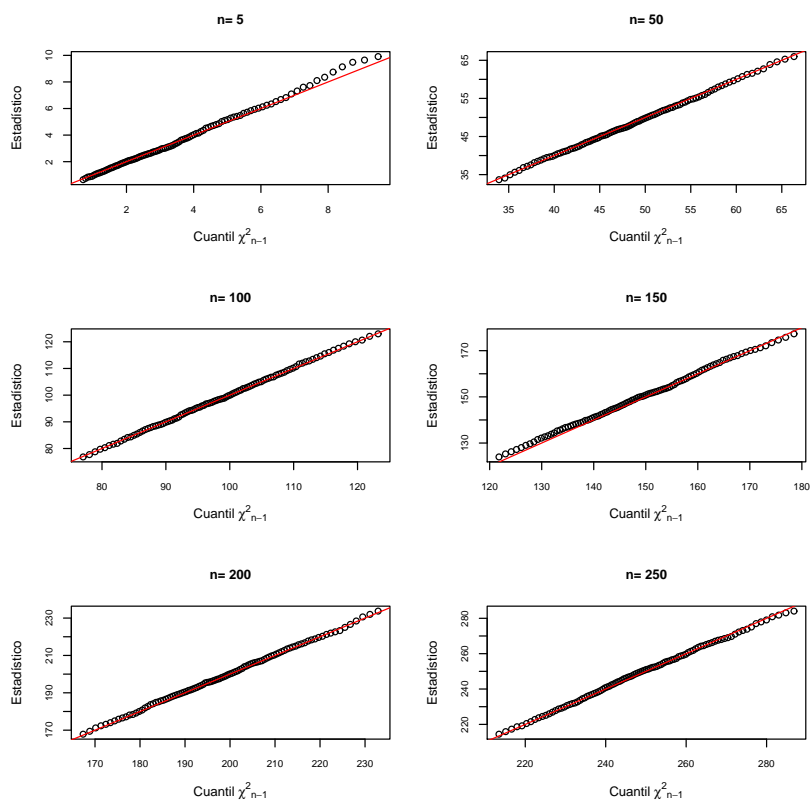


Figura 3.2: Gráficos QQ para cuantiles de la distribución probabilística  $\chi^2_{n-1}$  y cuantiles del estadístico  $(n-1)\widehat{\sigma}^2/\sigma^2$  basados en muestras seleccionadas de poblaciones infinitas. Los tamaños muestrales oscilan entre  $n = 5$  y  $n = 250$ .

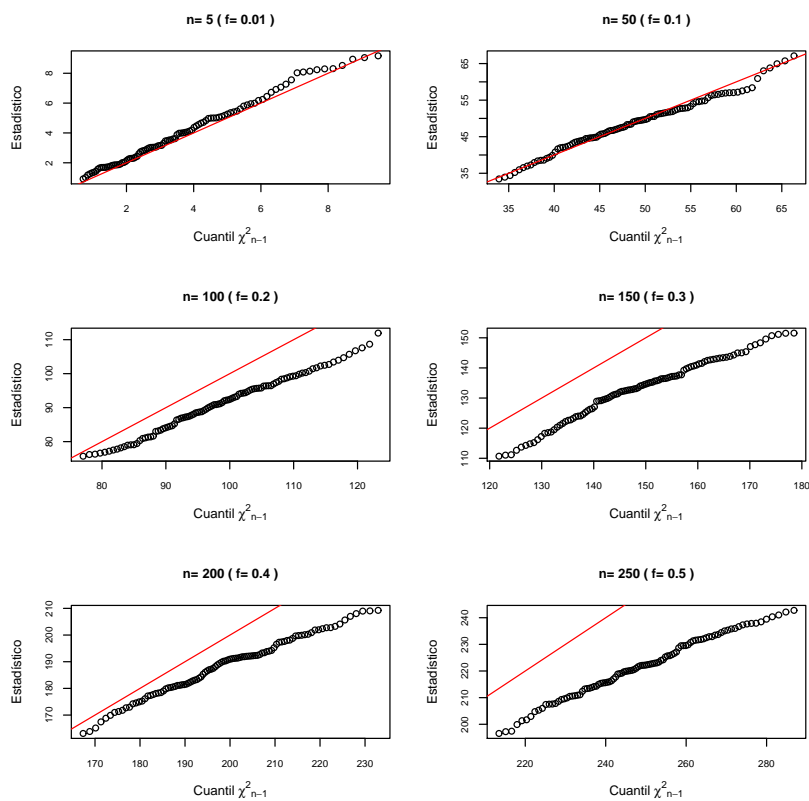


Figura 3.3: Gráficos QQ para cuantiles de la distribución probabilística  $\chi^2_{n-1}$  y cuantiles del estadístico  $(n-1)\widehat{\sigma}^2/\sigma^2$  basados en muestras seleccionadas de poblaciones finitas con fracciones de muestreo ( $f = n/N$ ) entre el 1% y el 50%. El tamaño de la población es  $N = 500$ , y los tamaños muestrales oscilan entre  $n = 5$  y  $n = 250$ .

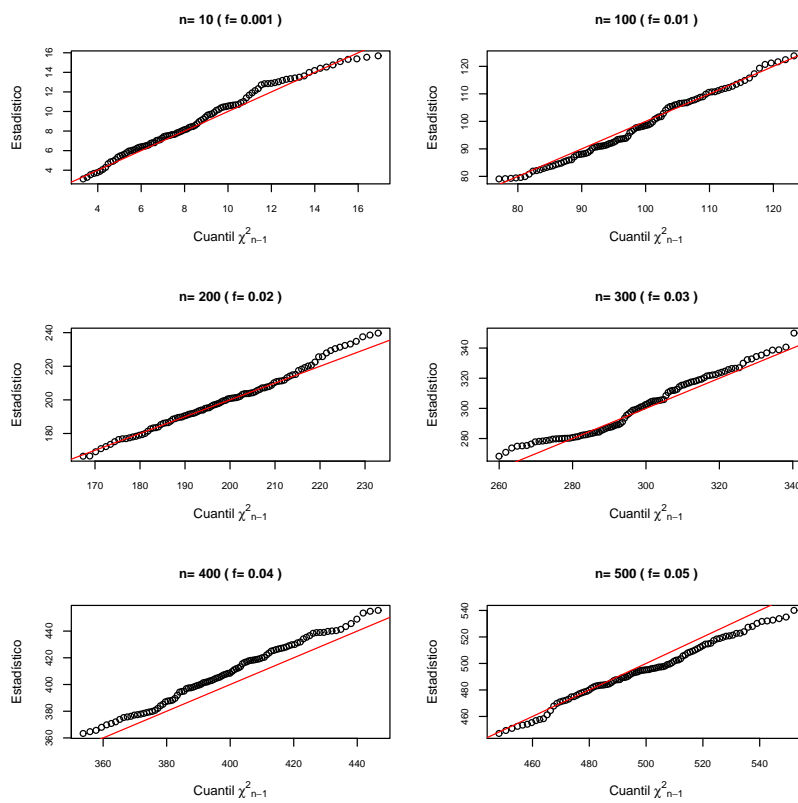


Figura 3.4: Gráficos QQ para cuantiles de la distribución probabilística  $\chi^2_{n-1}$  y cuantiles del estadístico  $(n-1)\widehat{\sigma}^2/\sigma^2$  basados en muestras seleccionadas de poblaciones finitas con fracciones de muestreo ( $f = n/N$ ) entre el 0.1 % y el 5 %. El tamaño de la población es  $N = 10000$ , y los tamaños muestrales oscilan entre  $n = 10$  y  $n = 500$ .

cantidad, se requiere el valor de  $c_4[498]$ , y así sucesivamente hasta llegar al menos a  $c_4[343]$ , que es el tamaño de muestra más elevado que se puede obtener de forma directa en R a partir de sus funciones existentes en este momento. Aunque esta operación es relativamente simple en el programa R, resultaría más conveniente el uso de una expresión directa para el cálculo de  $c_4[\cdot]$ . Por los motivos expuestos, se proponen expresiones que determinan el valor de  $c_4[\cdot]$  como productos de funciones, y evitando una excesiva recursividad. En la primera expresión propuesta se evalúa  $c_4[\cdot]$  en función de un único tamaño de muestra anterior. En la segunda expresión sugerida se calcula directamente el valor de  $c_4[\cdot]$  en el correspondiente tamaño de muestra, sin necesidad de evaluar la función en otro tamaño de muestra anterior.

En primer lugar, se incluyen en este trabajo una serie de resultados que son utilizados para la obtención de las nuevas expresiones para  $c_4[\cdot]$ . Por ejemplo, el Resultado 3.3. determina el valor  $c_4[n]$  cuando  $n = 2$ . Estos resultados están basados o hacen uso de las siguientes propiedades de la función  $\Gamma(\cdot)$ :

- (i).  $\Gamma(1) = 1$ ,
- (ii).  $\Gamma(1/2) = \sqrt{\pi}$ ,
- (iii).  $\Gamma(x + 1) = x\Gamma(x) \quad \forall x > 0$ .

**Resultado 3.3.** La función  $c_4[\cdot]$  evaluada en 2 tiene la siguiente expresión:

$$c_4[2] = \frac{\sqrt{2}}{\sqrt{\pi}}.$$

□

### Demostración

Siguiendo la definición de  $c_4[\cdot]$  y las propiedades de la función  $\Gamma(\cdot)$  anteriormente comentadas, nos queda:

$$c_4[2] = \frac{\sqrt{2}\Gamma(1)}{\sqrt{1}\Gamma(1/2)} = \frac{\sqrt{2}}{\sqrt{\pi}}.$$

□

**Resultado 3.4.** La función  $c_4[\cdot]$  evaluada en  $n + 1$  puede expresarse en función de la misma función  $c_4[\cdot]$ , pero evaluada en  $n$ , es decir:

$$c_4[n+1] = \frac{\sqrt{n-1}}{\sqrt{n} c_4[n]} \quad (3.3)$$

□

### Demostración

$$\begin{aligned} c_4[n+1] &= \frac{\sqrt{2}}{\sqrt{(n+1)-1}} \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{(n+1)-1}{2}\right)} = \frac{\sqrt{2}}{\sqrt{n}} \frac{\frac{n-1}{2} \Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = \\ &= \frac{\sqrt{2} \sqrt{2}}{\sqrt{2} \sqrt{n}} \frac{\frac{n-1}{2} \Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = \frac{(n-1)}{\sqrt{2} \sqrt{n}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = \\ &= \frac{\sqrt{n-1} \sqrt{n-1}}{\sqrt{2} \sqrt{n}} \frac{\Gamma\left(\frac{n-1}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} = \frac{\sqrt{n-1}}{\sqrt{n} \frac{\sqrt{2}}{\sqrt{n-1}} \Gamma\left(\frac{(n-1)}{2}\right)} = \\ &= \frac{\sqrt{n-1}}{\sqrt{n} c_4[n]}. \end{aligned}$$

□

La ecuación (3.3) se puede generalizar al caso  $n+k$  para un valor  $k > 1$ , tal como puede comprobarse en el siguiente resultado.

**Resultado 3.5.** La función  $c_4[\cdot]$  evaluada en  $n+k$ , con  $k \geq 1$  puede expresarse en función de la misma función  $c_4[\cdot]$ , pero evaluada en  $n+k-1$ , es decir:

$$c_4[n+k] = \frac{\sqrt{n+k-2}}{\sqrt{n+k-1} c_4[n+k-1]}$$

□

### Demostración



$$\begin{aligned}
c_4[n+k] &= \frac{\sqrt{2}}{\sqrt{(n+k)-1}} \frac{\Gamma\left(\frac{n+k}{2}\right)}{\Gamma\left(\frac{(n+k)-1}{2}\right)} = \\
&= \frac{\sqrt{2}}{\sqrt{n+k-1}} \frac{\frac{n+k-2}{2} \Gamma\left(\frac{n+k-2}{2}\right)}{\Gamma\left(\frac{n+k-1}{2}\right)} = \\
&= \frac{\sqrt{2}}{\sqrt{2}} \frac{\sqrt{2}}{\sqrt{n+k-1}} \frac{\frac{n+k-2}{2} \Gamma\left(\frac{n+k-2}{2}\right)}{\Gamma\left(\frac{n+k-1}{2}\right)} = \\
&= \frac{(n+k-2)}{\sqrt{2}\sqrt{n+k-1}} \frac{\Gamma\left(\frac{n+k-2}{2}\right)}{\Gamma\left(\frac{n+k-1}{2}\right)} = \\
&= \frac{\sqrt{n+k-2}\sqrt{n+k-2}}{\sqrt{2}\sqrt{n+k-1}} \frac{\Gamma\left(\frac{n+k-2}{2}\right)}{\Gamma\left(\frac{n+k-1}{2}\right)} = \\
&= \frac{\sqrt{n+k-2}}{\sqrt{n+k-1}} \frac{\sqrt{2}}{\sqrt{(n+k-1)-1}} \frac{\Gamma(n+k-1/2)}{\Gamma(((n+k-1)-1)/2)} = \\
&= \frac{\sqrt{n+k-2}}{\sqrt{n+k-1} c_4[n+k-1]}
\end{aligned}$$

□

La primera expresión propuesta para el cálculo de  $c_4[\cdot]$  puede consultarse en el Teorema 3.1. En este resultado se hace uso de las variables indicadoras  $I, I^c, I^*, J, J^c$  y  $J^*$ , las cuales se introducen para simplificar las expresiones resultantes, y se definen como sigue:

$$I = \begin{cases} 1 & \text{si } i \text{ es impar} \\ 0 & \text{si } i \text{ es par} \end{cases}$$

$$I^c = 1 - I = \begin{cases} 0 & \text{si } i \text{ es impar} \\ 1 & \text{si } i \text{ es par} \end{cases}$$

$$I^* = 2I - 1 = \begin{cases} 1 & \text{si } i \text{ es impar} \\ -1 & \text{si } i \text{ es par} \end{cases}$$

$$J = \begin{cases} 1 & \text{si } k \text{ es impar} \\ 0 & \text{si } k \text{ es par} \end{cases}$$

$$J^c = 1 - J = \begin{cases} 0 & \text{si } k \text{ es impar} \\ 1 & \text{si } k \text{ es par} \end{cases}$$

$$J^* = 2J - 1 = \begin{cases} 1 & \text{si } k \text{ es impar} \\ -1 & \text{si } k \text{ es par} \end{cases}$$

Los términos  $i$  y  $k$  también se introducen en el Teorema 3.1.

**Teorema 3.1.** Dado un tamaño de muestra  $n$  y un número entero  $k$ , donde  $n \geq 1$  y  $k \geq 3$ , la función  $c_4[\cdot]$  evaluada en  $n + k$  viene dada por

$$c_4[n + k] = \frac{(n - 1)^{J^*/2}(n + k - 2)}{(n + k - 1)^{1/2}(c_4[n])^{J^*}} \prod_{i=0}^{k-3} (n + i)^{I^* J^*} \quad \forall \quad k \geq 3 \quad (3.4)$$

□

### Demostración

Para demostrar la expresión (3.4) nos basamos en el conocido principio de inducción matemática (Franklin y Daoud 2011, Hermes 1973), el cual suele usarse para demostrar que ciertas propiedades matemáticas se verifican para todo número natural.

En primer lugar, se desarrolla la expresión del Resultado 3.5 para los primeros valores del término  $k$ . El objetivo de deducir una expresión para  $c_4[n + k]$ , es decir:

$$\begin{aligned}
c_4[n+1] &= \frac{\sqrt{(n+1)-2}}{\sqrt{(n+1)-1} c_4[(n+1)-1]} = \frac{\sqrt{n-1}}{\sqrt{n} c_4[n]} \\
c_4[n+2] &= \frac{\sqrt{(n+2)-2}}{\sqrt{(n+2)-1} c_4[(n+2)-1]} = \frac{\sqrt{n}}{\sqrt{n+1} c_4[n+1]} = \\
&= \frac{\sqrt{n}\sqrt{n} c_4[n]}{\sqrt{n+1}\sqrt{n-1}} = \frac{n c_4[n]}{\sqrt{n+1}\sqrt{n-1}} \\
c_4[n+3] &= \frac{\sqrt{(n+3)-2}}{\sqrt{(n+3)-1} c_4[(n+3)-1]} = \frac{\sqrt{n+1}}{\sqrt{n+2} c_4[n+2]} = \\
&= \frac{\sqrt{n+1}\sqrt{n+1}\sqrt{n-1}}{\sqrt{n+2} n c_4[n]} = \frac{(n+1)\sqrt{n-1}}{\sqrt{n+2} n c_4[n]}
\end{aligned}$$

A partir de las expresiones anteriores, supondremos que se verifica la siguiente relación

$$c_4[n+k] = \frac{(n-1)^{J^*/2} \prod_{i=0}^{k-2} (n+i)^{IJ+I^c J^c}}{(n+k-1)^{1/2} (c_4[n])^{J^*} \prod_{i=0}^{k-3} (n+i)^{I^c J+IJ^c}},$$

puesto que dependiendo de si los términos  $i$  y  $k$  son pares o impares, unas expresiones se encuentran en el numerador y otras en el denominador. Además, se puede observar que los dos términos producto se aplican para  $k \geq 3$ .

Supongamos que  $k$  es impar, el siguiente paso, siguiendo el principio de inducción matemática, sería comprobar que se verifica la relación propuesta para  $k^* = k + 1$ , donde  $k^*$  será por tanto un número par. Siguiendo estas indicaciones se tiene

$$\begin{aligned}
c_4[n+k^*] &= c_4[n+k+1] = \frac{\sqrt{n+k-1}}{\sqrt{n+k} c_4[n+k]} = \\
&= \frac{\sqrt{n+k-1}\sqrt{n+k-1}(n+k-3)\dots n c_4[n]}{\sqrt{n+k}(n+k-2)(n+k-4)\dots(n+1)\sqrt{n-1}} = \\
&= \frac{(n+k-1)(n+k-3)\dots n c_4[n]}{\sqrt{n+k}(n+k-2)(n+k-4)\dots(n+1)\sqrt{n-1}}
\end{aligned}$$

Sustituyendo  $k + 1 = k^*$  queda:

$$\begin{aligned}
& \frac{(n + k^* - 2)(n + k^* - 4) \dots n c_4[n]}{\sqrt{n + k^* - 1}(n + k^* - 3)(n + k^* - 5) \dots (n + 1)\sqrt{n - 1}} = \\
& \quad \frac{(n - 1)^{J^*/2} \prod_{i=0}^{k^*-2} (n + i)^{IJ + I^c J^c}}{(n + k^* - 1)^{1/2} (c_4[n])^{J^*} \prod_{i=0}^{k^*-3} (n + i)^{I^c J + I J^c}} = \\
& = \frac{(n - 1)^{J^*/2} (n + k^* - 2) \prod_{i=0}^{k^*-3} (n + i)^{I^* J^*}}{(n + k^* - 1)^{1/2} (c_4[n])^{J^*}} = c_4[n + k^*],
\end{aligned}$$

lo cual coincide con la expresión  $c_4[n + k]$  definida en (3.4), y después de sustituir  $k^*$  por  $k$ .

Tan sólo quedaría comprobar que la expresión propuesta es válida cuando  $k$  sea un número par, y por tanto, se hace la comprobación para el valor siguiente  $k^* = k + 1$ , donde  $k^*$  será impar:

$$\begin{aligned}
c_4[n + k^*] & = c_4[n + k + 1] = \frac{\sqrt{n + k - 1}}{\sqrt{n + k} c_4[n + k]} = \\
& = \frac{\sqrt{n + k - 1} \sqrt{n + k - 1} (n + k - 3) \dots (n + 1) \sqrt{n - 1}}{\sqrt{n + k} (n + k - 2) (n + k - 4) \dots n c_4[n]} = \\
& = \frac{(n + k - 1) (n + k - 3) \dots (n + 1) \sqrt{n - 1}}{\sqrt{n + k} (n + k - 2) (n + k - 4) \dots n c_4[n]}
\end{aligned}$$

A continuación, sustituyendo  $k + 1 = k^*$  se obtiene la siguiente expresión:

$$\begin{aligned}
& \frac{(n + k^* - 2)(n + k^* - 4) \dots (n + 1) \sqrt{n - 1}}{\sqrt{n + k^* - 1}(n + k^* - 3)(n + k^* - 5) \dots n c_4[n]} = \\
& \quad \frac{(n - 1)^{J^*/2} \prod_{i=0}^{k^*-2} (n + i)^{IJ + I^c J^c}}{(n + k^* - 1)^{1/2} (c_4[n])^{J^*} \prod_{i=0}^{k^*-3} (n + i)^{I^c J + I J^c}} = \\
& = \frac{(n - 1)^{J^*/2} (n + k^* - 2) \prod_{i=0}^{k^*-3} (n + i)^{I^* J^*}}{(n + k^* - 1)^{1/2} (c_4[n])^{J^*}} = c_4[n + k^*],
\end{aligned}$$

que también coincide con la expresión  $c_4[n + k]$  de la ecuación (3.4), y después de sustituir  $k^*$  por  $k$ .

En consecuencia, se concluye, en virtud del principio de inducción, que la expresión definida en la ecuación (3.4) es cierta para  $k \geq 3$ .  $\square$

**Teorema 3.2** Dado un tamaño de muestra  $n$ , con  $n \geq 2$ , la función  $c_4[\cdot]$  evaluada en  $n$  viene dada por

$$c_4[n] = \frac{(\pi/2)^{I_n/2} \prod_{i=2}^{n-2} i^{I_n I^*}}{(n - 1)^{1/2}} \quad (3.5)$$

donde  $I_n$  es una variable indicadora definida como:

$$I_n = \begin{cases} 1 & \text{si } n \text{ es impar} \\ -1 & \text{si } n \text{ es par} \end{cases}$$

y  $I_n I^*$  asume los siguientes valores:

$$I_n I^* = I_n (2I - 1) = \begin{cases} 1 & \begin{cases} \text{si } \{n, i\} \text{ es impar} \\ \text{si } \{n, i\} \text{ es par} \end{cases} \\ -1 & \begin{cases} \text{si } n \text{ es impar y } i \text{ es par} \\ \text{si } n \text{ es par y } i \text{ es impar} \end{cases} \end{cases}$$

$\square$

**Demostración**

Por una parte se demuestra que implementando el valor  $c_4[2] = \sqrt{2}/\sqrt{\pi}$  descrito en el Resultado 3.3. en la expresión (3.4) se obtiene:

$$c_4[2+k] = \begin{cases} \frac{k(k-2)(k-4)\cdots 3\sqrt{\pi}}{\sqrt{k+1}(k-1)(k-3)\cdots 2\sqrt{2}} & \text{si } k \text{ es impar} \\ \frac{k(k-2)(k-4)\cdots 2\sqrt{2}}{\sqrt{k+1}(k-1)(k-3)\cdots 3\sqrt{\pi}} & \text{si } k \text{ es par} \end{cases} \quad \forall k \geq 3$$

Por otra parte, desarrollando también la expresión (3.3) del Resultado 3.4 a partir del valor  $c_4[2] = \sqrt{2}/\sqrt{\pi}$ , se obtiene:

$$c_4[3] = \frac{\sqrt{\pi}}{\sqrt{2}\sqrt{2}} = \frac{\sqrt{\pi}}{2}$$

$$c_4[4] = \frac{\sqrt{2}\sqrt{2}\sqrt{2}}{\sqrt{3}\sqrt{\pi}} = \frac{(\sqrt{2})^3}{\sqrt{3}\sqrt{\pi}}$$

$$c_4[5] = \frac{\sqrt{3}\sqrt{3}\sqrt{\pi}}{\sqrt{4}\sqrt{2}\sqrt{2}\sqrt{2}} = \frac{3\sqrt{\pi}}{\sqrt{4}(\sqrt{2})^3}$$

$$c_4[6] = \frac{\sqrt{4}\sqrt{4}(\sqrt{2})^3}{\sqrt{5}3\sqrt{\pi}} = \frac{4(\sqrt{2})^3}{\sqrt{5}3\sqrt{\pi}}$$

Por tanto, haciendo uso de las variables indicadoras definidas con anterioridad y aplicando de nuevo el principio de inducción, se obtiene fácilmente el valor de esta función para un valor  $n$  par,

$$c_4[n] = \frac{\prod_{i=4}^{n-2} i^{I^c} 2^{3/2}}{\prod_{i=3}^{n-3} i^I \pi^{1/2} (n-1)^{1/2}}$$

Además, puesto que

$$\prod_{i=4}^{n-2} i^{I^c} = \prod_{i=3}^{n-2} i^{I^c}$$

puede simplificarse la expresión anterior dando lugar a la siguiente expresión:

$$c_4[n] = \frac{\prod_{i=2}^{n-2} i^{I^c} 2^{1/2}}{\prod_{i=2}^{n-2} i^I \pi^{1/2} (n-1)^{1/2}}$$

De forma análoga al caso de un tamaño de muestra par, puede obtenerse fácilmente la siguiente expresión para un valor  $n$  impar:

$$c_4[n] = \frac{\prod_{i=3}^{n-2} i^I \pi^{1/2}}{\prod_{i=4}^{n-2} i^{I^c} 2^{3/2} (n-1)^{1/2}} = \frac{\prod_{i=2}^{n-2} i^I \pi^{1/2}}{\prod_{i=2}^{n-2} i^{I^c} 2^{1/2} (n-1)^{1/2}}$$

En conclusión, para cualquier tamaño de muestra  $n$ , la función  $c_4[\cdot]$  puede expresarse como:

$$\begin{aligned} c_4[n] &= \frac{\pi^{I_n/2}}{(n-1)^{1/2} 2^{I_n/2}} \left( \prod_{i=2}^{n-2} (n-i)^{I^*} \right)^{I_n} = \\ &= \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_n I^*}. \end{aligned}$$

□

### 3.4. Conclusiones

En el contexto del SQC, el parámetro  $\sigma$  asociado a la desviación típica de una característica de calidad de un proceso se identifica con la variabilidad de

dicho proceso. Asimismo, la desviación típica juega un papel muy relevante en la materia de estudio principal de este trabajo: los índices de capacidad. Como se ha comentado en capítulos anteriores, los índices de capacidad comparan la amplitud de los límites de especificación con la variabilidad del proceso. En la práctica, el parámetro  $\sigma$  es desconocido y, por tanto, su estimación es también especialmente relevante y de gran interés en el campo del SQC.

El estimador insesgado de  $\sigma$  se ha denominado  $\hat{\sigma}_{c_4}$ , puesto que depende de la función  $c_4[\cdot]$ . La expresión matemática que define esta función presenta algunos inconvenientes, los cuales han sido discutidos a lo largo de este capítulo. Se han propuesto dos nuevas expresiones para  $c_4[\cdot]$ . Mediante la primera expresión, descrita en la ecuación (3.4) del Teorema 3.1, es posible calcular el valor de dicha función para un tamaño muestral concreto, y basándose en un valor de la propia función pero evaluada en otro tamaño de muestra inferior. Por otro lado, la segunda expresión, descrita en la ecuación (3.5), permite el cálculo directo de  $c_4[\cdot]$ .

Las ventajas de las nuevas expresiones se resumen a continuación. En primer lugar, se ha ganado en simplicidad. En este sentido, las nuevas expresiones definen  $c_4[\cdot]$  como un producto de fracciones. En segundo lugar, las expresiones propuestas permiten el cálculo de la función cuando el tamaño de la muestra es elevado, a diferencia de la expresión tradicional que obtiene infinitos a través de la función  $\Gamma(\cdot)$ . En otras palabras, las nuevas expresiones permiten la disponibilidad del valor de  $c_4[\cdot]$  para cualquier tamaño de muestra. Destacamos que este hecho es transcendental en este trabajo. Por ejemplo, en el Capítulo 2, buena parte del análisis del índice de capacidad se efectuaba asumiendo que este indicador empresarial es estimado basándose en una única muestra. Además, se señaló que esta opción es muy común en la práctica, puesto que si la empresa tiene la certeza de que el proceso está bajo control, podría suponer un menor coste que su alternativa, en la cual se seleccionan múltiples muestras a lo largo del tiempo en lugar de una única muestra. Así pues, es razonable suponer que al tratarse de una única muestra, ésta debe ser de un tamaño considerable. Por último, el hecho de contar con expresiones más simples y capaces de obtener el valor de la función para cualquier tamaño muestral, hace posible programar en un programa estadístico una función que determine su valor y, así, poder automatizar el control de calidad de la empresa. En este sentido, se ha elaborado la función `CONST.C4` en el software estadístico R que obtiene el valor de  $c_4[\cdot]$  para cualquier tamaño muestral. Esta función se incluye en la librería “Capability” que se presenta en el Capítulo 4 de este trabajo.

El estudio de la insesgades de estimadores de la desviación típica poblacional tiene su base en la distribución chi-cuadrado, es decir, a través del siguiente



estadístico

$$Y = \frac{(n-1)\widehat{\sigma}^2}{\sigma^2} \rightsquigarrow \chi_{n-1}^2.$$

A lo largo de este trabajo se han realizado comparaciones entre poblaciones finitas e infinitas, y este estudio también resulta interesante para analizar el cumplimiento de dicha distribución bajo ambos tipos de poblaciones. Por ejemplo, el incumplimiento de esta distribución puede tener un impacto negativo en la calidad de intervalos de confianza para la varianza poblacional, los cuales están basados en esta distribución. En este sentido, se ha realizado un estudio basado en gráficos QQ, y en el cual se compara la distribución empírica del estadístico  $Y$  con la correspondiente distribución teórica. Se han considerado muestras seleccionadas de poblaciones finitas e infinitas. En el caso de población finita se han diferenciado fracciones de muestreo elevadas y próximas a cero. Se observa que, en el caso de población infinita, la distribución empírica se aproxima a la distribución teórica. En caso de población finita, la fracción de muestreo considerada es relevante, ya que a medida que se aumenta ésta la distribución empírica del estadístico se distancia más a la teórica. De este modo, se observa un buen ajuste entre las distribuciones empíricas y teóricas en el caso de fracciones de muestreo próximas a cero. Por el contrario, considerando fracciones de muestreo elevadas la distancia entre ambas distribuciones puede ser significativa.

# Capítulo 4

## A new R-package: ‘Capability’

### 4.1. Introduction. The statistical R software

The main aim of this research is to provide new contributions and analysis related to the estimation of capability indexes of a production process, which are the parameters of interest in this study. The idea was to analyze the performance of various estimators of capability indexes under different possible scenarios that could appear in practice. A summary of the various contributions of this study is as follows.

First, an extensive revision of the literature related to the SQC, and the capability index in particular, has been carried out. The estimation of the standard deviation is an important issue in our research, since the parameters of interest depend on the standard deviation. Second, we analyzed the empirical performance of estimators of the standard deviation and the effect of them on the estimation of capability indexes. Third, we also analyzed the estimation of the standard deviation and the capability indexes when the process is out of control. It is quite common to assume infinite populations in the context of SQC. Fourth, we compared the performance of estimators under both finite and infinite populations. Fifth, we analyzed the distribution of the sample variance via some QQ-plots. Sixth, estimators were also analyzed under the presence of populations without normality. Seventh, some of the previous contribution were extended to the case of  $m$  samples, where  $m > 1$ . The constant  $c_4$  is very common in the problem of estimating capability indexes, since this constant provides unbiased estimators of the population standard deviation. Finally, we have also proposed new expressions for the constant  $c_4$ , and which have the

advantage of being calculable for any sample size.

The development of the mentioned contributions have required the use of specific software. We considered the statistical software *R* (R Core Team 2016) because of many reasons. For instance, *R* is a constantly devolving program. In fact, every approximately six months a new version of the program is presented. As mentioned in Section 1.3, *R* is composed of a basic library, which contains basic functions to work with *R* and other eight libraries that serve as support. A library or package is set of *R* objects that deal with specific topics. Many additional libraries about different topics are also available. The new versions of *R* improve certain aspects of the base library. In addition, any user can develop a library, and it may be included in the list of available libraries. We emphasize that *R* is a free software, multisystem and capable of performing high quality graphics, so it is increasingly used both in terms of research and in business. One of the most important advantages of *R* is its programming language, which is flexible and powerful. It contains a multitude of commands and allows us to create our own functions and routines. In particular, *R* is the appropriate tool to perform Monte Carlo simulation studies. In addition, the execution of the programmed functions is very fast and efficient, hence results can be quickly obtained and well organized in a file for reading. For the various mentioned reasons, *R* is a tool with a wide variety of applications in many fields. In particular, *R* has special relevance in social sciences.

The present study required a revision of the different packages included in *R* and related to SQC. In particular, the *R*-packages “AcceptanceSamplig”, “IQCC”, “qcc”, “qualityTools”, “SixSigma” and “spc” were analyzed.

The library “AcceptanceSamplig” focuses on acceptance sampling. It provides functionality for creating and evaluating acceptance sampling plans. Sampling plans can be single, double or multiple.

The library “IQCC” focuses on control chart. This library builds statistical control charts with exact limits for univariate and multivariate cases.

The library “qcc” provides functions to plot the standard quality control charts for various types of data (continuous, attribute and count data), the corresponding operating characteristic curves, the Cumsum and EWMA charts for continuous data, and the Pareto chart. In addition, “qcc” plots the cause-and-effect diagram and a basic process capability analysis.

The library “qualityTools” contains methods associated with the Define, Measure, Analyze, Improve and Control (i.e. DMAIC) cycle of the Six Sigma Quality Management methodology. It covers distribution fitting, normal and

non-normal process capability indices, techniques for Measurement Systems Analysis especially gage capability indices and Gage Repeatability (i.e Gage RR) and Reproducibility studies, factorial and fractional factorial designs as well as response surface methods including the use of desirability functions. Improvement via Six Sigma is project based strategy that covers 5 phases: Define - Pareto Chart; Measure - Probability and Quantile- Quantile Plots, Process Capability Indices for various distributions and Gage RR Analyze i.e. Pareto Chart, Multi-Vari Chart, Dot Plot; Improve - Full and fractional factorial, response surface and mixture designs as well as the desirability approach for simultaneous optimization of more than one response variable. Normal, Pareto and Lenth Plot of effects as well as Interaction Plots; Control - Quality Control Charts can be found in the 'qcc' package. The focus is on teaching the statistical methodology used in the Quality Sciences.

The library "SixSigma" provides functions and utilities to perform Statistical Analyses in the Six Sigma way. Through the DMAIC cycle (Define, Measure, Analyze, Improve, Control), we can manage several Quality Management studies: Gage R & R, Capability Analysis, Control Charts, Loss Function Analysis, etc. Data frames used in Cano et al. (2012) and Cano et al. (2015) are also included in this package.

The library "spc" evaluation of control charts by means of the zero-state, steady-state ARL (Average Run Length) and RL quantiles. Setting up control charts for given in-control ARL. The control charts under consideration are one- and two-sided EWMA, CUSUM, and Shiryaev-Roberts schemes for monitoring the mean of normally distributed independent data. ARL calculation of the same set of schemes under drift are added. Other charts and parameters are in preparation. Further SPC areas will be covered as well (sampling plans, capability indices, etc.).

Note that none of the existing *R* packages include a deep study related to the process capability analysis, hence the construction of a *R* package focuses on the process capability analysis may have a relevant interest for researches and companies. The aim of the proposed *R* package, which is named as "Capability", is to provide the real process capability indexes when the corresponding parameters are known. If the parameters are unknown, the *R* package "Capability" computes different point estimators of the capability indexes and the corresponding confidence intervals based on different methodologies.

## 4.2. The package: ‘Capability’

**Title** Process Capability Analysis

**Description** Functions to calculate and estimate Process Capability Indexes (*PCIs*). The *PCIs* depend on the population (true) process standard deviation. Two different cases are considered. First, the true *PCIs* are calculated when the parameters associated to the process are known. Second, different point estimators and confidence intervals are computed when the parameters are unknown.

### R topics documented

APROX.FISHER	132
APROX.WH	134
CI.Cp.F.Sc4	135
CI.Cp.Heavlin.Sc4	138
CI.Cp.Sc4	141
CI.Cp.WH.Sc4	143
CI.Cpk.Sc4	146
CONST.C4	149
CONST.D2	151
Cp	152
Cp.M.R	153
Cp.M.R.WM	155
Cp.M.Sc4	158
Cp.M.Sp	160
Cp.Sc4	163
Cpk	165
Cpk.M.R	167
Cpk.M.R.WM	170
Cpk.M.Sc4	172
Cpk.M.Sp	176
Cpk.Sc4	179
Cpl	182
Cpl.Sc4	183
Cpu	186
Cpu.Sc4	187

SIGMA.M.R	.....	190
SIGMA.M.R.WM	.....	191
SIGMA.M.Sc4	.....	193
SIGMA.M.Sp	.....	196
SIGMA.Sc4	.....	198

## Index

# Description

---

Capability-package

*Process Capability Analysis package*

---

## Description

The package contains functions to calculate and estimate the most common Process Capability Indexes (*PCIs*) related to a process capability analysis. The *PCIs* depend on the process standard deviation. Two different cases are considered. First, the true *PCIs* are calculated when the parameters associated to the process are known. Second, different point estimators and confidence intervals are computed when the parameters are unknown. Estimation methods are based on the sample standard deviation and the sample ranges. Moreover, this package contains functions to derive different confidence intervals for the various *PCIs*. Additional functions related to the process capacity analysis are also included.

The whole list of available functions are listed below. They are grouped in lists, aiming to clarify their usage.

### Known parameters:

Cp  
Cpk  
Cpu  
Cpl

**Estimates based on a single sample:**

SIGMA.S  
Cp.Sc4  
Cpk.Sc4  
Cpu.Sc4  
Cpl.Sc4  
CI.Cp.Sc4  
CI.Cp.F.Sc4  
CI.Cp.WH.Sc4  
CI.Cp.Heavlin.Sc4  
CI.Cpk.Sc4

**Estimates based on  $m$  samples:**

SIGMA.M.R  
SIGMA.M.R.WM  
SIGMA.M.Sc4  
SIGMA.M.Sp  
Cp.M.R  
Cp.M.R.WM  
Cp.M.Sc4  
Cp.M.Sp  
Cpk.M.R  
Cpk.M.R.WM  
Cpk.M.Sc4  
Cpk.M.Sp

**Complementary functions:**

APROX.FISHER  
APROX.WH  
CONST.C4  
CONST.D2

## Functions

---

APROX.FISHER	Fisher's approximation to the square root of the chi-square distribution
--------------	--

---

### Description

Fisher's approximation to the square root of the chi-square distribution with  $df$  degrees of freedom.

### Usage

APROX.FISHER(a,df)

### Arguments

a	vector of probabilities.
df	degrees of freedom (non-negative, but can be non-integer).

### Details

The Fisher's approximation to the square root of the chi-square distribution with  $df$  degrees of freedom is defined as:

$$\chi_{df,a} \cong \sqrt{df - \frac{1}{2}} + \frac{Z_a}{\sqrt{2}}$$

where  $Z_a$  is the  $a$ th quantile of the standard normal distribution.

It is quite common to replace the quantile used in the construction of confidence intervals for a given *PCI* by valuable approximations suggested by Fisher (1928).

### Value

The function returns the value of the  $a$ th quantile of the Fisher's approximation to the square root of the chi-square distribution with  $df$  degrees of freedom.

### References

Fisher, R. A. (1928). *Statistical Methods for Research Workers*, 2nd Edition, pp. 96-7.

Wilson, E. B., Hilferty, M. M. (1931). The distribution of chi-square. *Proceedings of the National Academy of Sciences*, 17(12): 684-688.

### See also

APROX.WH



qchisq

### Examples

APROX.FISHER(0.95, 10)

APROX.FISHER(0.975, 10)

APROX.FISHER(0.995, 10)

---

APROX.WH	Wilson-Hirferty's approximation to the square root of the chi-square distribution
----------	---

---

### Description

The Wilson-Hirferty's approximation to the square root of the chi-square distribution with  $df$  degrees of freedom.

### Usage

APROX.WH(a,df)

### Arguments

a                      vector of probabilities.  
df                     degrees of freedom (non-negative, but can be non-integer).

### Details

The Wilson-Hirferty's approximation to the square root of the chi-square distribution with  $df$  degrees of freedom is defined as:

$$\chi_{df,a} \cong df^{\frac{1}{2}} \left[ 1 - \frac{2}{9df} + Z_a \left( \frac{2}{9df} \right)^{\frac{1}{2}} \right]^{\frac{3}{2}}$$

where  $Z_a$  is the  $a$ th quantile of the standard normal distribution.

It is quite common to replace the quantile used in the construction of confidence intervals for a given *PCI* by valuable approximations suggested by Wilson-Hirferty (1931).

## Value

The function returns the value of the  $a$ th quantile of the Wilson-Hilferty's approximation to the square root of the chi-square distribution with  $df$  degrees of freedom.

## References

Wilson, E. B., Hilferty, M. M. (1931). The distribution of chi-square. *Proceedings of the National Academy of Sciences*, 17(12): 684-688.

Fisher, R. A. (1928). *Statistical Methods for Research Workers*, 2nd Edition, pp. 96-7.

## See also

APROX.FISHER

qchisq

## Examples

APROX.WH(0.95, 10)

APROX.WH(0.975, 10)

APROX.WH(0.995, 10)

---

CI.Cp.F.Sc4	Confidence Interval for a Process Capability Index using the Fisher's approximation
-------------	---

---

## Description

Computes the confidence interval of the Process Capability Index based on the sample standard deviation. The Fisher's approximation to the square root of the chi-square distribution is considered.

## Usage

CI.Cp.F.Sc4(Xn, LSL, USL, CL)

## Arguments

$X_n$	vector of the quality characteristic of interest $X$ ; its length is equal to $n$ .
LSL	the lower specification limit.
USL	the upper specification limit.
CL	the confidence level $(1 - \alpha)$ for the confidence interval.

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and a single sample with size  $n$ , the point estimator of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the sample standard deviation is defined as:

$$\widehat{C}_{p.c4} = \frac{LSE - LIE}{6\hat{\sigma}_{c4}},$$

where

$$\hat{\sigma}_{c4} = \frac{S}{c_4[n]}$$

is the unbiased estimator of the process standard deviation based on the sample standard deviation,

$$S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2},$$

the sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

is the estimator of the process mean,  $x_i$  is the value of the quality characteristic  $X$  for the  $i$ th product, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Finally,  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

The  $(1 - \alpha)\%$  confidence interval of the Process Capability Index  $C_p$  using the Fisher's approximation is defined as:

$$L = \frac{1}{\sqrt{n-1}} \hat{C}_{p.c4} \left( \sqrt{n - \frac{3}{2}} - \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{2}} \right)$$

and

$$U = \frac{1}{\sqrt{n-1}} \hat{C}_{p.c4} \left( \sqrt{n - \frac{3}{2}} + \frac{Z_{1-\frac{\alpha}{2}}}{\sqrt{2}} \right)$$

where  $L$  and  $U$  are the lower and upper limits of the interval respectively, and  $Z_{1-\frac{\alpha}{2}}$  is the  $1 - \alpha/2$ th quantile of the standard Normal distribution.

## Value

The function returns the confidence interval for the Process Capability Index using the Fisher's approximation.

## References

Kotz, S., Johnson, N. L. (1993). *Process capability indices*. CRC Press.

## See also

CI.Cp.Heavlin.Sc4

CI.Cp.Sc4

CI.Cp.WH.Sc4

CI.Cpk.Sc4

## Examples

```
DATA1 <- rnorm(100, 10, 1)
```

```
CI.Cp.F.Sc4 (DATA1, 7, 13, 0.95)
```

```
DATA2 <- rnorm(100, 10, 0.5)
```

```
CI.Cp.F.Sc4 (DATA2, 7, 13, 0.95)
```

```
DATA3 <- rnorm(100, 10, 2)
```

```
CI.Cp.F.Sc4 (DATA3, 7, 13, 0.95)
```

---

CI.Cp.Heavlin.Sc4	Confidence Interval for a Process Capability Index suggested by Heavlin (1988)
-------------------	--

---

## Description

Computes the confidence interval of the Process Capability Index suggested by Heavlin (1988) and based on the sample standard deviation.

## Usage

```
CI.Cp.Heavlin.Sc4(Xn, LSL, USL, CL)
```

## Arguments

Xn	vector of the quality characteristic of interest $X$ ; its length is equal to $n$ .
LSL	the lower specification limit.
USL	the upper specification limit.
CL	the confidence level $(1 - \alpha)$ for the confidence interval.

## Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and a single sample with size  $n$ , the point estimator of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the sample standard deviation is defined as:

$$\widehat{C}_{p.c4} = \frac{LSE - LIE}{6\hat{\sigma}_{c4}},$$

where

$$\hat{\sigma}_{c4} = \frac{S}{c_4[n]}$$

is the unbiased estimator of the process standard deviation based on the sample standard deviation,

$$S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2},$$

the sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

is the estimator of the process mean,  $x_i$  is the value of the quality characteristic  $X$  for the  $i$ th product, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Finally,  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

The  $(1 - \alpha)$  % confidence interval of the Process Capability Index  $C_p$  suggested by Heavlin (1988) is defined as  $[L, U]$ , where:

$$L = \widehat{C}_{p.c4} \left( 1 - \left( \frac{1}{2(n-3)} \left( 1 + \frac{6}{n-1} \right) \right)^{\frac{1}{2}} Z_{1-\alpha/2} \right)$$

and

$$U = \widehat{C}_{p.c4} \left( 1 + \left( \frac{1}{2(n-3)} \left( 1 + \frac{6}{n-1} \right) \right)^{\frac{1}{2}} Z_{1-\alpha/2} \right)$$

where  $Z_{1-\frac{\alpha}{2}}$  is the  $1 - \alpha/2$ th quantile of the standard Normal distribution.

### Value

The function returns the confidence interval for the Process Capability Index suggested by Heavlin (1988).

### References

Heavlin, W.D. (1988). Statistical properties of capability indices. *Technical Report 320*

Kotz, S., Johnson, N. L. (1993). *Process capability indices*. CRC Press.

### See also

CI.Cp.F.Sc4

CI.Cp.Sc4

CI.Cp.WH.Sc4

CI.Cpk.Sc4

### Examples

```
DATA1 <- rnorm(100, 10, 1)
```

```
CI.Cp.Heavlin.Sc4(DATA1, 7, 13, 0.95)
```

```
DATA2 <- rnorm(100, 10, 0.5)
```

```
CI.Cp.Heavlin.Sc4(DATA2, 7, 13, 0.95)
```

DATA3 < -rnorm(100, 10, 2)

CI.Cp.Heavlin.Sc4(DATA3, 7, 13, 0.95)

---

CI.Cp.Sc4	The customary confidence Interval for a Process Capability Index
-----------	--

---

### Description

Computes the customary confidence interval of the Process Capability Index based on the sample standard deviation.

### Usage

CI.Cp.Sc4(Xn, LSL, USL, CL)

### Arguments

Xn	vector of the quality characteristic of interest $X$ ; its length is equal to $n$ .
LSL	the lower specification limit.
USL	the upper specification limit.
CL	the confidence level $(1 - \alpha)$ for the confidence interval.

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and a single sample with size  $n$ , the point estimator of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the sample standard deviation is defined as:

$$\hat{C}_{p.c4} = \frac{LSE - LIE}{6\hat{\sigma}_{c4}},$$

where

$$\hat{\sigma}_{c4} = \frac{S}{c_4[n]}$$



is the unbiased estimator of the process standard deviation based on the sample standard deviation,

$$S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2},$$

the sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

is the estimator of the process mean,  $x_i$  is the value of the quality characteristic  $X$  for the  $i$ th product, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Finally,  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

The  $(1 - \alpha)$  % confidence interval of the Process Capability Index is defined by:

$$\left[ \frac{\chi_{n-1, \alpha/2}}{\sqrt{n-1}} \widehat{C}_{p.c4}, \frac{\chi_{n-1, 1-\alpha/2}}{\sqrt{n-1}} \widehat{C}_{p.c4} \right],$$

where  $\chi_{n-1,1-\alpha/2}$  is the square root of the  $1 - \alpha/2$ th quantile of the chi-square distribution.

### Value

The function returns the customary confidence interval for the Process Capability Index.

### References

Kotz, S., Johnson, N. L. (1993). *Process capability indices*. CRC Press.

### See also

CI.Cp.F.Sc4

CI.Cp.Heavlin.Sc4

CI.Cp.WH.Sc4

CI.Cpk.Sc4

### Examples

```
DATA1 <- rnorm(100, 10, 1)
```

```
CI.Cp.Sc4 (DATA1, 7, 13, 0.95)
```

```
DATA2 <- rnorm(100, 10, 0.5)
```

```
CI.Cp.Sc4 (DATA2, 7, 13, 0.95)
```

```
DATA3 <- rnorm(100, 10, 2)
```

```
CI.Cp.Sc4 (DATA3, 7, 13, 0.95)
```

---

CI.Cp.WH.Sc4	Confidence Interval for a Process Capability Index using the Wilson-Hirferty's approximation
--------------	--

---

### Description

Computes the confidence interval of the Process Capability Index based on the sample standard deviation. The Wilson-Hirferty's approximation to the square root of the chi-square distribution is considered.

## Usage

CI.Cp.WH.Sc4(Xn, LSL, USL, CL)

## Arguments

Xn	vector of the quality characteristic of interest $X$ ; its length is equal to $n$ .
LSL	the lower specification limits.
USL	the upper specification limits.
CL	the confidence level $(1 - \alpha)$ for the confidence interval.

## Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and a single sample with size  $n$ , the point estimator of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the sample standard deviation is defined as:

$$\hat{C}_{p.c4} = \frac{LSE - LIE}{6\hat{\sigma}_{c4}},$$

where

$$\hat{\sigma}_{c4} = \frac{S}{c_4[n]}$$

is the unbiased estimator of the process standard deviation based on the sample standard deviation,

$$S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2},$$

the sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

is the estimator of the process mean,  $x_i$  is the value of the quality characteristic  $X$  for the  $i$ th product, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_n^i}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Finally,  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

The  $(1 - \alpha)\%$  confidence interval of the Process Capability Index using the Wilson-Hirferty's approximation is defined as:

$$L = \widehat{C}_{p.c4} \left( 1 - \frac{2}{9(n-1)} - Z_{1-\frac{\alpha}{2}} \sqrt{\frac{2}{9(n-1)}} \right)$$

and

$$U = \widehat{C}_{p.c4} \left( 1 - \frac{2}{9(n-1)} + Z_{1-\frac{\alpha}{2}} \sqrt{\frac{2}{9(n-1)}} \right),$$

where  $L$  and  $U$  are the lower and upper limits of the interval, respectively, and  $Z_{1-\frac{\alpha}{2}}$  is the  $1 - \alpha/2$ th quantile of the standard Normal distribution.

### Value

The function returns the confidence interval for the Process Capability Index using the Wilson-Hirferty's approximation.

### References

Kotz, S., Johnson, N. L. (1993). *Process capability indices*. CRC Press.

### See also

CI.Cp.Heavlin.Sc4

CI.Cp.Sc4

CI.Cp.F.Sc4

CI.Cpk.Sc4

### Examples

DATA1 <- rnorm(100, 10, 1)

CI.Cp.WH.Sc4(DATA1, 7, 13, 0.95)

DATA2 <- rnorm(100, 10, 0.5)

CI.Cp.WH.Sc4(DATA2, 7, 13, 0.95)

DATA3 <- rnorm(100, 10, 2)

CI.Cp.WH.Sc4(DATA3, 7, 13, 0.95)

---

CI.Cpk.Sc4	Confidence Interval for a Process Capability Index for an Off-Center Process
------------	--

---

### Description

Computes the confidence interval of the Process Capability Index for an Off-Center Process and using the sample standard deviation.

### Usage

CI.Cpk.Sc4(Xn, LSL, USL, CL, MU)

### Arguments

Xn	vector of the quality characteristic of interest $X$ ; its length is equal to $n$ .
LSL	the lower specification limits.
USL	the upper specification limits.
CL	the confidence level $(1 - \alpha)$ for the confidence interval.
MU	the process mean. If the process mean is not indicated, the function uses the sample mean, i.e. the mean of Xn.

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and a single sample with size  $n$ , the point estimator of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the sample standard deviation is defined as:

$$\widehat{C}_{p.c4} = \frac{LSE - LIE}{6\hat{\sigma}_{c4}},$$

where

$$\hat{\sigma}_{c4} = \frac{S}{c_4[n]}$$

is the unbiased estimator of the process standard deviation based on the sample standard deviation,

$$S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2},$$

the sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

is the estimator of the process mean,  $x_i$  is the value of the quality characteristic  $X$  for the  $i$ th product, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Finally,  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

A process is Off-Center when

$$k = \frac{|\mu_0 - \mu|}{(LSE - LIE)/2} \neq 0$$

i.e., the relative distance between the process mean,  $\mu$ , and the midpoint of the specification limits,

$$\mu_0 = \frac{LSE + LIE}{2},$$

is not equal to 0. The customary point estimator of the Process Capability Index for an Off-Center Process,

$$C_{pk} = (1 - k)C_p$$

and using the sample standard deviation is defined as

$$\widehat{C}_{pk.c4} = (1 - \widehat{k})\widehat{C}_{p.c4},$$

where

$$\widehat{k} = \frac{|\mu_0 - \widehat{\mu}|}{(LSE - LIE)/2}.$$

It is common to use the sample mean when the parameter  $\mu$  is unknown, i.e.,  $\widehat{\mu} = \bar{x}$ . If the argument MU is given, this function uses the real value of  $k$  instead of  $\widehat{k}$ .

The  $(1 - \alpha)\%$  confidence interval of the Process Capability Index for an Off-Center Process is given by  $[L_k, U_k]$ , where:

$$L_k = \widehat{C}_{pk.c4} - Z_{1-\frac{\alpha}{2}} \left( \frac{n-1}{9n(n-3)} + \widehat{C}_{pk.c4}^2 \frac{1}{2(n-3)} \left( 1 + \frac{6}{n-1} \right) \right)^{\frac{1}{2}}$$

and

$$U_k = \widehat{C}_{pk.c4} + Z_{1-\frac{\alpha}{2}} \left( \frac{n-1}{9n(n-3)} + \widehat{C}_{pk.c4}^2 \frac{1}{2(n-3)} \left( 1 + \frac{6}{n-1} \right) \right)^{\frac{1}{2}},$$

where  $Z_{1-\frac{\alpha}{2}}$  is the  $1 - \alpha/2$ th quantile of the standard Normal distribution.

### Value

The function returns the confidence interval of the Process Capability Index for an Off-Center process.

### References

Kotz, S., Johnson, N. L. (1993). *Process capability indices*. CRC Press.

### See also

CI.Cp.F.Sc4

CI.Cp.Heavlin.Sc4

CI.Cp.WH.Sc4

CI.Cpk.Sc4

### Examples

```
MU <- -10
```

```
DATA1 <- rnorm(100, MU, 1)
```

```
CI.Cpk.Sc4(DATA1, 7, 13, 0.95, MU)
```

```
MU <- -11
```

```
DATA2 <- rnorm(100, MU, 1)
```

```
CI.Cpk.Sc4(DATA2, 7, 13, 0.95, MU)
```

```
DATA3 <- rnorm(100, 10, 1)
```

```
CI.Cpk.Sc4(DATA3, 7, 13, 0.95)
```

---

CONST.C4	Constant $c_4$ for a given sample size
----------	--

---

### Description

Calculates the value of the constant  $c_4$  for a given sample size.

### Usage



CONST.C4( $n$ )

### Arguments

$n$  non-negative integer of length 1 that indicates the sample size.

### Details

For a given sample size  $n$ , the constant  $c_4$  is used to correct the bias of the sample standard deviation, i.e., the usual estimator of the population standard deviation. The corresponding values of  $c_4[\cdot]$ , for any  $n$ , are given by:

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

### Value

The function returns the value of the constant  $c_4$  for the sample size  $n$ .

### References

Chapter 3 of the recent document.

### See also

CONST.D2

SIGMA.Sc4

Cp.Sc4

### Examples

CONST.C4(5)

CONST.C4(10)

CONST.C4(400)

---

CONST.D2	Constant $d_2$ for a given sample size
----------	--

---

### Description

Calculates the value of the constant  $d_2$  for a given sample size.

### Usage

CONST.D2(n)

### Arguments

n                      non-negative integer of length 1 that indicates the sample size.

### Details

For a given sample size  $n$ , the constant  $d_2$  is used to compute the estimator of the population standard deviation based on sample relative ranges. For a given  $n$ , the values of  $d_2[\cdot]$  is defined as:

$$d_2[n] = \int_{-\infty}^{\infty} 1 - (Z_y)^n - (1 - Z_y)^n dy,$$

where  $Z_y$  is the probability associated with the  $y$ th quantile of the standard normal distribution.

### Value

The function returns the value of the constant  $d_2$  for a given sample size  $n$ .

### References

Mitra A. (2008). *Fundamentals of quality control and improvement* Hoboken, N.J. Wiley.

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

**See also**

CONST.C4

**Examples**

CONST.D2(5)

CONST.D2(10)

CONST.D2(400)

---

Cp	Process Capability Index of a centered process
----	--

---

**Description**

Calculates the true value of the Process Capability Index of a centered process.

**Usage**

Cp(SIGMA, LSL, USL)

**Arguments**

SIGMA	positive number that indicates the value of the process standard deviation.
LSL	the lower specification limit.
USL	the upper specification limit.

**Details**

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process, the Process Capability Index compares the natural variability of the process with the amplitude of the specification limits, i.e.

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

where  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. If  $C_p = 1$ , the process is marginally suitable to meet the specification limits; if  $C_p > 1$ , the process is suitable to meet the specification limits; and if  $C_p < 1$ , the process is not able to meet the specification limits.

### Value

The function returns the true value of the Process Capability Index of a centered process.

### References

Kotz, S., Johnson, N. L. (1993). *Process capability indices*. CRC Press.

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

### See also

Cpu

Cpl

Cpk

Cp.Sc4

### Examples

Cp(1, 7, 13)

Cp(0.5, 7, 13)

Cp(2, 7, 13)

---

Cp.M.R	Estimator of a Process Capability Index for a centered process and based on the sample ranges derived from $m$ samples
--------	--

---

### Description

Computes the estimator of the Process Capability Index for a centered process and using the sample ranges derived from  $m$  samples.

### Usage

Cp.M.R(Xn, LSL, USL)

### Arguments

Xn	matrix (with dimension $m \times n$ ) of the quality characteristic of interest $X$ ; each row of Xn corresponds to a sample with length equal to $n$ .
LSL	the lower specification limit.
USL	the upper specification limit.

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and  $m$  samples with the same size  $n$ , the point estimator of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the sample ranges is defined as:

$$\hat{C}_{p.R} = \frac{USL - LSL}{6\hat{\sigma}_R},$$

where

$$\hat{\sigma}_R = \frac{\bar{R}}{d_2[n]}$$

is the estimator of the process standard deviation based on the sample ranges

$$R_i = \text{máx}(x_{ij}) - \text{mín}(x_{ij}),$$

$x_{ij}$  denotes the observed value of  $X$  for the  $j$ th product, with  $j = 1, \dots, n$ , in the  $i$ th sample, with  $i = 1, \dots, m$ ,

$$\bar{R} = \frac{1}{m} \sum_{i=1}^m R_i$$

is the average of the sample ranges,  $d_2[n]$  is defined as

$$d_2[n] = \int_{-\infty}^{\infty} 1 - (Z_y)^n - (1 - Z_y)^n dy,$$

and  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $d_2[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009).

### Value

The function returns the value of the point estimator of the Process Capability Index of a centered process and using the sample ranges derived from  $m$  samples.

### References

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

Woodall W.H., Montgomery D.C. (2000). Using Ranges to Estimate Variability. *Quality Engineering* ; **13(2)**: 211-217.

### See also

Cp.M.R.WM

Cp.M.Sc4

Cp.M.Sp

### Examples

```
DATA1 <- matrix(rnorm(100, 0, 1), nc=20)
```

```
Cp.M.R(DATA1, 7, 13)
```

```
DATA2 <- matrix(rnorm(20, 10, 2), nc=4)
```

```
Cp.M.R(DATA2, 7, 13)
```

```
DATA3 <- matrix(rnorm(100, 0, 2), nc=5)
```

```
Cp.M.R(DATA3, -3, 3)
```

---

Cp.M.R.WM	Estimator based on the sample ranges defined by Woodall and Montgomery (2000) of a Process Capability Index of a centered process
-----------	---

---

### Description

Estimates of Process Capability Index of a centered process using the point estimator based on the sample ranges defined by Woodall and Montgomery (2000).

### Usage

Cp.M.R.WM(Xn)

### Arguments

Xn	matrix (with dimension $m \times n$ ) of the quality characteristic of interest $X$ ; each row of Xn corresponds to a sample with length equal to $n$ .
LSL	the lower specification limit.
USL	the upper specification limit.

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and  $m$  samples with the same size  $n$ , the point estimator defined by Woodall and Montgomery (2000) of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the sample ranges is defined as:

$$\hat{C}_{p.R.WM} = \frac{USL - LSL}{6\hat{\sigma}_{R.WM}},$$

where

$$\hat{\sigma}_{R.WM} = \frac{d_2[n]}{d_2^*[n]}\bar{R}$$

is the estimator of the process standard deviation based on the sample ranges

$$R_i = \text{máx}(x_{ij}) - \text{mín}(x_{ij}),$$

$x_{ij}$  denotes the observed value of  $X$  for the  $j$ th product, with  $j = 1, \dots, n$ , in the  $i$ th sample, with  $i = 1, \dots, m$ ,

$$\bar{R} = \frac{1}{m} \sum_{i=1}^m R_i$$

is the average of the sample ranges,  $d_2[n]$  is defined as

$$d_2[n] = \int_{-\infty}^{\infty} 1 - (Z_y)^n - (1 - Z_y)^n dy,$$

$d_2^*[n]$  is defined as

$$d_2^*[n] = \left( d_2^2[n] + \frac{d_3^2[n]}{m} \right)^{1/2},$$

$d_3[n]$  is defined by Woodall and Montgomery (2000), and  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $d_2[n]$  and  $d_3[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009).

### Value

The function returns the value of the point estimator defined by Woodall and Montgomery (2000) of the Process Capability Index of a centered process and using the sample ranges derived from  $m$  samples.

### References

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

Woodall W.H., Montgomery D.C. (2000). Using Ranges to Estimate Variability. *Quality Engineering* ; **13(2)**: 211-217.

### See also

Cp.M.R

Cp.M.Sc4

Cp.M.Sp

### Examples

```
DATA1 <- matrix(rnorm(100, 10, 1), nc=20)
```

```
Cp.M.R.WM(DATA1, 7, 13)
```

```
DATA2 <- matrix(rnorm(20, 10, 2), nc=4)
```

```
Cp.M.R.WM(DATA2, 7, 13)
```

```
DATA3 <- matrix(rnorm(100, 0, 2), nc=5)
```



Cp.M.R.WM(DATA3, -3, 3)

---

Cp.M.Sc4	Estimator of a Process Capability Index for a centered process and based on the sample standard deviations derived from $m$ samples
----------	---

---

### Description

Computes the estimator of the Process Capability Index for a centered process and using the sample standard deviations derived from  $m$  samples.

### Usage

Cp.M.Sc4(Xn, LSL, USL)

### Arguments

Xn	matrix (with dimension $m \times n$ ) of the quality characteristic of interest $X$ ; each row of Xn corresponds to a sample with length equal to $n$ .
LSL	the lower specification limit.
USL	the upper specification limit.

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and  $m$  samples with the same size  $n$ , the point estimator of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the sample standard deviations is defined as:

$$\hat{C}_{p.S} = \frac{USL - LSL}{6\hat{\sigma}_S}$$

where

$$\hat{\sigma}_S = \frac{\bar{S}}{c_4[n]}$$

is the estimator of the process standard deviation based on the sample standard deviations, where

$$\bar{S} = \frac{1}{m} \sum_{i=1}^m S_i$$

is the average of the sample standard deviations

$$S_i = \left( \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \right)^{1/2},$$

$x_{ij}$  denotes the observed value of  $X$  for the  $j$ th product, with  $j = 1, \dots, n$ , in the  $i$ th sample, with  $i = 1, \dots, m$ ,

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$$

is the sample mean for the  $i$ th sample, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Finally,  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

## Value

The function returns the value of the point estimator of the Process Capability Index for a centered process and using the sample standard deviations derived from  $m$  samples.

## References

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

Chen G. (1997). The mean and standard deviation of the run length distribution of X-bar charts when control limits are estimated. *Statistica Sinica* 7: 789-798.

### See also

Cp.M.R

Cp.M.R.WM

Cp.M.Sp

### Examples

```
DATA1 <- matrix(rnorm(100, 10, 1), nc=20)
```

```
Cp.M.Sc4(DATA1, 7, 13)
```

```
DATA2 <- matrix(rnorm(100, 10, 2), nc=4)
```

```
Cp.M.Sc4(DATA2, 7, 13)
```

```
DATA3 <- matrix(rnorm(100, 0, 2), nc=5)
```

```
Cp.M.Sc4(DATA3, -3, 3)
```

---

Cp.M.Sp	Estimator of a Process Capability Index for a centered process and based on the pooled sample standard deviation obtained from $m$ samples
---------	--

---

### Description

Computes the estimator of the Process Capability Index for a centered process and using the pooled sample standard deviation obtained from  $m$  samples.

### Usage

```
Cp.M.Sp(Xn, LSL, USL)
```

### Arguments

Xn	matrix (with dimension $m \times n$ ) of the quality characteristic of interest $X$ ; each row of Xn corresponds to a sample with length equal to $n$ .
LSL	the lower specification limit.
USL	the upper specification limit.

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and  $m$  samples with the same size  $n$ , the point estimator of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the pooled sample standard deviation is defined as:

$$\hat{C}_{p.Sp} = \frac{USL - LSL}{6\hat{\sigma}_{Sp}},$$

where

$$\hat{\sigma}_{Sp} = \frac{S_p}{c_4[1 + m(n - 1)]},$$

is the estimator of the process standard deviation based on the sample standard deviations

$$S_i = \left( \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \right)^{1/2},$$

$x_{ij}$  denotes the observed value of  $X$  for the  $j$ th product, with  $j = 1, \dots, n$ , in the  $i$ th sample, with  $i = 1, \dots, m$ ,

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$$

is the sample mean for the  $i$ th sample,

$$S_p = \left( \frac{1}{m} \sum_{i=1}^m S_i^2 \right)^{1/2}$$

is the pooled sample standard deviation, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_n^i}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Finally,  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

## Value

The function returns the value of the point estimator of the Process Capability Index for a centered process and using the pooled sample standard deviation obtained from  $m$  samples.

## References

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

Chen G. (1997). The mean and standard deviation of the run length distribution of X-bar charts when control limits are estimated. *Statistica Sinica* 7: 789-798.

## See also

Cp.M.R

Cp.M.R.WM

Cp.M.Sc4

## Examples

```
DATA1 <- matrix(rnorm(100, 10, 1), nc=20)
```

Cp.M.Sp(DATA1, 7, 13)

DATA2 <- matrix(rnorm(100, 10, 2), nc=4)

Cp.M.Sp(DATA2, 7, 13)

DATA3 <- matrix(rnorm(100, 0, 2), nc=5)

Cp.M.Sp(DATA3, -3, 3)

---

Cp.Sc4	Estimator of a Process Capability Index for a centered process and based on the sample standard deviation
--------	---

---

### Description

Computes the estimator of the Process Capability Index for a centered process and using the sample standard deviation.

### Usage

Cp.Sc4(Xn, LSL, USL)

### Arguments

Xn                      vector of the quality characteristic of interest  $X$ ; its length is equal to  $n$ .  
LSL                      the lower specification limit.  
USL                      the upper specification limit.

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and a single sample with size  $n$ , the point estimator of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the sample standard deviation is defined as:

$$\hat{C}_{p.c4} = \frac{LSE - LIE}{6\hat{\sigma}_{c4}},$$

where

$$\hat{\sigma}_{c_4} = \frac{S}{c_4[n]}$$

is the unbiased estimator of the process standard deviation based on the sample standard deviation,

$$S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2},$$

the sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

is the estimator of the process mean,  $x_i$  is the value of the quality characteristic  $X$  for the  $i$ th product, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Finally,  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

## Value

The function returns the value of the point estimator of the Process Capability Index for a centered process and using the sample standard deviation.

## References

Kotz, S., Johnson, N. L. (1993). *Process capability indices*. CRC Press.

## See also

Cpk.Sc4

Cpu.Sc4

Cpl.Sc4

## Examples

```
DATA1 <- rnorm(100, 10, 1)
```

```
Cp.Sc4(DATA1, 7, 13)
```

```
DATA2 <- rnorm(100, 10, 0.5)
```

```
Cp.Sc4(DATA2, 7, 13)
```

```
DATA3 <- rnorm(100, 10, 2)
```

```
Cp.Sc4(DATA3, 7, 13)
```

---

Cpk	Process Capability Index of an off-center process
-----	---

---

## Description

Calculates the value of the Process Capability Index of an off-center process.

## Usage

```
Cpk(MU, SIGMA, LSL, USL)
```

## Arguments

MU	the process mean.
SIGMA	a positive number that indicate the process standard deviation.
LSL	the lower specification limit.
USL	the upper specification limit.



## Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process, the Process Capability Index compares the natural variability of the process with the amplitude of the specification limits, i.e.

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

where  $LSL$  and  $USL$  are the lower and upper specification limits, respectively.

A process is Off-Center when

$$k = \frac{|\mu_0 - \mu|}{(LSE - LIE)/2} \neq 0$$

i.e., the relative distance between the process mean,  $\mu$ , and the midpoint of the specification limits,

$$\mu_0 = \frac{LSE + LIE}{2},$$

is not equal to 0. The Process Capability Index for an Off-Center Process is defined as

$$C_{pk} = (1 - k)C_p$$

where

$$\hat{k} = \frac{|\mu_0 - \mu|}{(LSE - LIE)/2}.$$

## Value

The function returns the true value of the Process Capability Index of an off-center process.

## References

Kotz, S., Johnson, N. L. (1993). *Process capability indices*. CRC Press.

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

## See also

Cpu

Cpl

Cp

Cp.c4

### Examples

Cpk(10, 1, 7, 13)

Cpk(11, 1, 7, 13)

Cpk(11, 0.5, 7, 13)

---

Cpk.M.R	Estimator of a Process Capability Index for an off-centered process and based on the sample ranges derived from $m$ samples
---------	---

---

### Description

Computes the estimator of the Process Capability Index for an off-centered process and using the sample ranges derived from  $m$  samples.

### Usage

Cpk.M.R(Xn, LSL, USL, MU)

### Arguments

Xn	matrix (with dimension $m \times n$ ) of the quality characteristic of interest $X$ ; each row of Xn corresponds to a sample with length equal to $n$ .
LSL	the lower specification limit.
USL	the upper specification limit.
MU	the process mean. If the process mean is not indicated, the function uses the sample mean, i.e. the mean of Xn.

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and  $m$  samples with the same size  $n$ , the point estimator of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the sample ranges is defined as:

$$\widehat{C}_{p.R} = \frac{USL - LSL}{6\widehat{\sigma}_R},$$

where

$$\widehat{\sigma}_R = \frac{\overline{R}}{d_2[n]}$$

is the estimator of the process standard deviation based on the sample ranges

$$R_i = \max(x_{ij}) - \min(x_{ij}),$$

$x_{ij}$  denotes the observed value of  $X$  for the  $j$ th product, with  $j = 1, \dots, n$ , in the  $i$ th sample, with  $i = 1, \dots, m$ ,

$$\overline{R} = \frac{1}{m} \sum_{i=1}^m R_i$$

is the average of the sample ranges,  $d_2[n]$  is defined as

$$d_2[n] = \int_{-\infty}^{\infty} 1 - (Z_y)^n - (1 - Z_y)^n dy,$$

and  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $d_2[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009).

A process is Off-Center when

$$k = \frac{|\mu_0 - \mu|}{(LSE - LIE)/2} \neq 0$$

i.e., the relative distance between the process mean,  $\mu$ , and the midpoint of the specification limits,

$$\mu_0 = \frac{LSE + LIE}{2},$$

is not equal to 0. The customary point estimator of the Process Capability Index for an Off-Center Process,

$$C_{pk} = (1 - k)C_p$$

and using the sample standard deviation is defined as

$$\widehat{C}_{pk.R} = (1 - \widehat{k})\widehat{C}_{p.R},$$

where

$$\hat{k} = \frac{|\mu_0 - \hat{\mu}|}{(LSE - LIE)/2}.$$

It is common to use the sample mean when the parameter  $\mu$  is unknown, i.e.,  $\hat{\mu} = \bar{x}$ . If the argument MU is given, this function uses the real value of  $k$  instead of  $\hat{k}$ .

### Value

The function returns the value of the point estimator of the Process Capability Index of an off-centered process and using the sample ranges derived from  $m$  samples.

### References

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

Woodall W.H., Montgomery D.C. (2000). Using Ranges to Estimate Variability. *Quality Engineering* ; **13(2)**: 211-217.

### See also

Cpk.M.R.WM

Cpk.M.Sc4

Cpk.M.Sp

### Examples

```
MU <- - 10
```

```
DATA1 <- matrix(rnorm(100,MU, 1), nc=20)
```

```
Cpk.M.R(DATA1, 7, 13, MU)
```

```
MU <- - 11
```

```
DATA2 <- matrix(rnorm(100, MU, 2), nc=4)
```

```
Cpk.M.R(DATA2, 7, 13, MU)
```

```
DATA3 <- matrix(rnorm(100, 10, 1), nc=5)
```

```
Cpk.M.R(DATA3, 7, 13)
```

---

Cpk.M.R.WM	Estimator based on the sample ranges defined by Woodall and Montgomery (2000) of a Process Capability Index of an off-centered process
------------	--

---

## Description

Estimates of Process Capability Index of an off-centered process using the point estimator based on the sample ranges defined by Woodall and Montgomery (2000).

## Usage

Cpk.M.R.WM(Xn)

## Arguments

Xn	matrix (with dimension $m \times n$ ) of the quality characteristic of interest $X$ ; each row of Xn corresponds to a sample with length equal to $n$ .
LSL	the lower specification limit.
USL	the upper specification limit.
MU	the process mean. If the process mean is not indicated, the function uses the sample mean, i.e. the mean of Xn.

## Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and  $m$  samples with the same size  $n$ , the point estimator defined by Woodall and Montgomery (2000) of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the sample ranges is defined as:

$$\hat{C}_{p.R.WM} = \frac{USL - LSL}{6\hat{\sigma}_{R.WM}},$$

where

$$\hat{\sigma}_{R.WM} = \frac{d_2[n]}{d_2^*[n]}\bar{R}$$

is the estimator of the process standard deviation based on the sample ranges

$$R_i = \max(x_{ij}) - \min(x_{ij}),$$

$x_{ij}$  denotes the observed value of  $X$  for the  $j$ th product, with  $j = 1, \dots, n$ , in the  $i$ th sample, with  $i = 1, \dots, m$ ,

$$\bar{R} = \frac{1}{m} \sum_{i=1}^m R_i$$

is the average of the sample ranges,  $d_2[n]$  is defined as

$$d_2[n] = \int_{-\infty}^{\infty} 1 - (Z_y)^n - (1 - Z_y)^n dy,$$

$d_2^*[n]$  is defined as

$$d_2^*[n] = \left( d_2^2[n] + \frac{d_3^2[n]}{m} \right)^{1/2},$$

$d_3[n]$  is defined by Woodall and Montgomery (2000), and  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $d_2[n]$  and  $d_3[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009).

A process is Off-Center when

$$k = \frac{|\mu_0 - \mu|}{(LSE - LIE)/2} \neq 0$$

i.e., the relative distance between the process mean,  $\mu$ , and the midpoint of the specification limits,

$$\mu_0 = \frac{LSE + LIE}{2},$$

is not equal to 0. The customary point estimator of the Process Capability Index for an Off-Center Process,

$$C_{pk} = (1 - k)C_p$$

and using the sample standard deviation is defined as

$$\hat{C}_{pk.RWM} = (1 - \hat{k})\hat{C}_{p.RWM},$$

where

$$\hat{k} = \frac{|\mu_0 - \hat{\mu}|}{(LSE - LIE)/2}.$$

It is common to use the sample mean when the parameter  $\mu$  is unknown, i.e.,  $\hat{\mu} = \bar{x}$ . If the argument MU is given, this function uses the real value of  $k$  instead of  $\hat{k}$ .

## Value

The function returns the value of the point estimator defined by Woodall and Montgomery (2000) of the Process Capability Index of an off-centered process and using the sample ranges derived from  $m$  samples.

## References

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

Woodall W.H., Montgomery D.C. (2000). Using Ranges to Estimate Variability. *Quality Engineering* ; **13(2)**: 211-217.

## See also

Cpk.M.R

Cpk.M.Sc4

Cpk.M.Sp

## Examples

```
MU <- - 10
```

```
DATA1 <- matrix(rnorm(100, MU, 1), nc=20)
```

```
Cpk.M.R.WM(DATA1, 7, 13, MU)
```

```
MU <- - 11
```

```
DATA2 <- matrix(rnorm(100, MU, 2), nc=4)
```

```
Cpk.M.R.WM(DATA2, 7, 13, MU)
```

```
DATA3 <- matrix(rnorm(100, 10, 1), nc=5)
```

```
Cpk.M.R.WM(DATA3, 7, 13)
```

---

Cpk.M.Sc4	Estimator of a Process Capability Index for an off-centered process and based on the sample standard deviations derived from $m$ samples
-----------	--

---

## Description

Computes the estimator of the Process Capability Index for an off-centered process and using the sample standard deviations derived from  $m$  samples.

### Usage

Cpk.M.Sc4(Xn, LSL, USL, MU)

### Arguments

Xn	matrix (with dimension $m \times n$ ) of the quality characteristic of interest $X$ ; each row of Xn corresponds to a sample with length equal to $n$ .
LSL	the lower specification limit.
USL	the upper specification limit.
MU	the process mean. If the process mean is not indicated, the function uses the sample mean, i.e. the mean of Xn.

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and  $m$  samples with the same size  $n$ , the point estimator of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the sample standard deviations is defined as:

$$\hat{C}_{p.S} = \frac{USL - LSL}{6\hat{\sigma}_S}.$$

where

$$\hat{\sigma}_S = \frac{\bar{S}}{c_4[n]}$$

is the estimator of the process standard deviation based on the sample standard deviations, where

$$\bar{S} = \frac{1}{m} \sum_{i=1}^m S_i$$

is the average of the sample standard deviations

$$S_i = \left( \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \right)^{1/2},$$



$x_{ij}$  denotes the observed value of  $X$  for the  $j$ th product, with  $j = 1, \dots, n$ , in the  $i$ th sample, with  $i = 1, \dots, m$ ,

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$$

is the sample mean for the  $i$ th sample, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Finally,  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

A process is Off-Center when

$$k = \frac{|\mu_0 - \mu|}{(LSE - LIE)/2} \neq 0$$

i.e., the relative distance between the process mean,  $\mu$ , and the midpoint of the specification limits,

$$\mu_0 = \frac{LSE + LIE}{2},$$

is not equal to 0. The customary point estimator of the Process Capability Index for an Off-Center Process,

$$C_{pk} = (1 - k)C_p$$

and using the sample standard deviation is defined as

$$\widehat{C}_{pk.S} = (1 - \widehat{k})\widehat{C}_{p.S},$$

where

$$\widehat{k} = \frac{|\mu_0 - \widehat{\mu}|}{(LSE - LIE)/2}.$$

It is common to use the sample mean when the parameter  $\mu$  is unknown, i.e.,  $\widehat{\mu} = \bar{x}$ . If the argument MU is given, this function uses the real value of  $k$  instead of  $\widehat{k}$ .

### Value

The function returns the value of the point estimator of the Process Capability Index for an off-centered process and using the sample standard deviations derived from  $m$  samples.

### References

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

Chen G. (1997). The mean and standard deviation of the run length distribution of X-bar charts when control limits are estimated. *Statistica Sinica* 7: 789-798.

### See also

Cpk.M.R

Cpk.M.R.WM

Cpk.M.Sp

### Examples

```
MU <- -10
```

```
DATA1 <- matrix(rnorm(100, MU, 1), nc=20)
```

```
Cpk.M.Sc4(DATA1, 7, 13, MU)
```

```
MU <- -11
```

```
DATA2 <- matrix(rnorm(100, MU, 2), nc=4)
```

```
Cpk.M.Sc4(DATA2, 7, 13, MU)
```

```
DATA3< -matrix(rnorm(100, 10, 1), nc=5)
```

```
Cpk.M.Sc4(DATA3, 7, 13)
```

---

Cpk.M.Sp	Estimator of a Process Capability Index for an off-centered process and based on the pooled sample standard deviation obtained from $m$ samples
----------	---

---

### Description

Computes the estimator of the Process Capability Index for an off-centered process and using the pooled sample standard deviation obtained from  $m$  samples.

### Usage

```
Cpk.M.Sp(Xn, LSL, USL, MU)
```

### Arguments

Xn	matrix (with dimension $m \times n$ ) of the quality characteristic of interest $X$ ; each row of Xn corresponds to a sample with length equal to $n$ .
LSL	the lower specification limit.
USL	the upper specification limit.
MU	the process mean. If the process mean is not indicated, the function uses the sample mean, i.e. the mean of Xn.

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and  $m$  samples with the same size  $n$ , the point estimator of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the pooled sample standard deviation is defined as:

$$\hat{C}_{p.Sp} = \frac{USL - LSL}{6\hat{\sigma}_{Sp}},$$

where

$$\hat{\sigma}_{Sp} = \frac{S_p}{c_4[1 + m(n - 1)]},$$

is the estimator of the process standard deviation based on the sample standard deviations

$$S_i = \left( \frac{1}{n - 1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \right)^{1/2},$$

$x_{ij}$  denotes the observed value of  $X$  for the  $j$ th product, with  $j = 1, \dots, n$ , in the  $i$ th sample, with  $i = 1, \dots, m$ ,

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$$

is the sample mean for the  $i$ th sample,

$$S_p = \left( \frac{1}{m} \sum_{i=1}^m S_i^2 \right)^{1/2}$$

is the pooled sample standard deviation, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n - 1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Finally,  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated

(see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

A process is Off-Center when

$$k = \frac{|\mu_0 - \mu|}{(LSE - LIE)/2} \neq 0$$

i.e., the relative distance between the process mean,  $\mu$ , and the midpoint of the specification limits,

$$\mu_0 = \frac{LSE + LIE}{2},$$

is not equal to 0. The customary point estimator of the Process Capability Index for an Off-Center Process,

$$C_{pk} = (1 - k)C_p$$

and using the sample standard deviation is defined as

$$\widehat{C}_{pk.Sp} = (1 - \widehat{k})\widehat{C}_{p.Sp},$$

where

$$\widehat{k} = \frac{|\mu_0 - \widehat{\mu}|}{(LSE - LIE)/2}.$$

It is common to use the sample mean when the parameter  $\mu$  is unknown, i.e.,  $\widehat{\mu} = \bar{x}$ . If the argument MU is given, this function uses the real value of  $k$  instead of  $\widehat{k}$ .

## Value

The function returns the value of the point estimator of the Process Capability Index for an off-centered process and using the pooled sample standard deviation obtained from  $m$  samples.

## References

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

Chen G. (1997). The mean and standard deviation of the run length distribution of X-bar charts when control limits are estimated. *Statistica Sinica* 7: 789-798.

## See also

Cpk.M.R

Cpk.M.R.WM

Cpk.M.Sc4

### Examples

MU < - 10

DATA1 < -matrix(rnorm(100,MU, 1), nc=20)

Cpk.M.Sp(DATA1, 7, 13, MU)

MU < - 11

DATA2 < -matrix(rnorm(100, MU, 2), nc=4)

Cpk.M.Sp(DATA2, 7, 13, MU)

DATA3 < -matrix(rnorm(100, 10, 1), nc=5)

Cpk.M.Sp(DATA3, 7, 13)

---

Cpk.Sc4	Estimator of a Process Capability Index for an off-centered process and based on the sample standard deviation
---------	--

---

### Description

Computes the estimator of the Process Capability Index for an off-centered process and using the sample standard deviation.

### Usage

Cpk.Sc4(Xn, LSL, USL, MU)

### Arguments

Xn	vector of the quality characteristic of interest $X$ ; its length is equal to $n$ .
LSL	the lower specification limit.
USL	the upper specification limit.
MU	the process mean. If the process mean is not indicated, the function uses the sample mean, i.e. the mean of Xn.

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and a single sample with size  $n$ , the point estimator of the Process Capability Index

$$C_p = \frac{LSE - LIE}{6\sigma}$$

based on the sample standard deviation is defined as:

$$\widehat{C}_{p.c4} = \frac{LSE - LIE}{6\hat{\sigma}_{c4}},$$

where

$$\hat{\sigma}_{c4} = \frac{S}{c_4[n]}$$

is the unbiased estimator of the process standard deviation based on the sample standard deviation,

$$S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2},$$

the sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

is the estimator of the process mean,  $x_i$  is the value of the quality characteristic  $X$  for the  $i$ th product, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Finally,  $LSL$  and  $USL$  are the lower and upper specification limits, respectively. Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

A process is Off-Center when

$$k = \frac{|\mu_0 - \mu|}{(LSE - LIE)/2} \neq 0$$

i.e., the relative distance between the process mean,  $\mu$ , and the midpoint of the specification limits,

$$\mu_0 = \frac{LSE + LIE}{2},$$

is not equal to 0. The customary point estimator of the Process Capability Index for an Off-Center Process,

$$C_{pk} = (1 - k)C_p$$

and using the sample standard deviation is defined as

$$\widehat{C}_{pk.Sc4} = (1 - \widehat{k})\widehat{C}_{p.Sc4},$$

where

$$\widehat{k} = \frac{|\mu_0 - \widehat{\mu}|}{(LSE - LIE)/2}.$$

It is common to use the sample mean when the parameter  $\mu$  is unknown, i.e.,  $\widehat{\mu} = \bar{x}$ . If the argument MU is given, this function uses the real value of  $k$  instead of  $\widehat{k}$ .

## Value

The function returns the value of the point estimator of the Process Capability Index for a centered process and using the sample standard deviation.

## References

Kotz, S., Johnson, N. L. (1993). *Process capability indices*. CRC Press.

## See also

Cp.Sc4

Cpu.Sc4

Cpl.Sc4



## Examples

```
MU < -10
```

```
DATA1 < -rnorm(100, MU, 1)
```

```
Cpk.Sc4(DATA1, 7, 13, MU)
```

```
MU < -11
```

```
DATA2 < -rnorm(100, MU, 0.5)
```

```
Cpk.Sc4(DATA2, 7, 13, MU)
```

```
DATA3 < -rnorm(100, 10, 2)
```

```
Cpk.Sc4(DATA3, 7, 13)
```

---

Cpl	One-sided Process Capability Index (Lower specification only)
-----	---

---

## Description

Calculates the true value of the one-sided Process Capability Index (Lower specification only).

## Usage

```
Cpl(LSL, MU, SIGMA)
```

## Arguments

LSL	the lower specification limit.
MU	the process mean.
SIGMA	a positive number that indicate the process standard deviation.

## Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . The one-sided Process Capability Index (lower specification only) is defined as:

$$C_{pl} = \frac{\mu - LSL}{3\sigma},$$

where  $LSL$  is the lower specification limit.

### Value

The function returns the true value of the one-sided Process Capability Index (lower specification only).

### References

Kotz, S., Johnson, N. L. (1993). *Process capability indices*. CRC Press.

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

### See also

Cp

Cpk

Cpu

### Examples

Cpl(7, 10, 1)

Cpl(7, 10, 0.5)

Cpl(7, 10, 2)

---

Cpl.Sc4	Estimator of one-sided Process Capability Index (Lower specification only) based on the sample standard deviation
---------	---

---

### Description

Computes the estimator of one-sided Process Capability Index (lower specification only) using the point estimator based on the sample standard deviation.

### Usage

Cpl.Sc4(Xn, LSL, MU)

### Arguments

Xn	vector of the quality characteristic of interest $X$ ; its length is equal to $n$ .
LSL	the lower specification limit.
MU	the process mean. If the process mean is not indicated, the function uses the sample mean, i.e. the mean of Xn.

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and a single sample with size  $n$ , the point estimator of the one-side Process Capability Index (upper specification only)

$$C_{pl} = \frac{\mu - LSL}{3\sigma},$$

based on the sample standard deviation is defined as:

$$\hat{C}_{pl.c4} = \frac{\mu - LSL}{3\hat{\sigma}_{c4}},$$

where

$$\hat{\sigma}_{c4} = \frac{S}{c_4[n]}$$

is the unbiased estimator of the process standard deviation based on the sample standard deviation,

$$S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2},$$

the sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

is the estimator of the process mean,  $x_i$  is the value of the quality characteristic  $X$  for the  $i$ th product, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_{n/2}}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{n,i}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Finally,  $LSL$  is the lower specification limit. Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

It is common to use the sample mean ( $\bar{x}$ ) when the parameter  $\mu$  is unknown, i.e.,  $\hat{\mu} = \bar{x}$ .

## Value

The function returns the value of the point estimator of the one-sided Process Capability Index (lower specification only) using the sample standard deviation.

## References

Kotz, S., Johnson, N. L. (1993). *Process capability indices*. CRC Press.

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

## See also

Cpl

Cpu.Sc4

## Examples

```
MU < -10
```

```
DATA1 < -rnorm(100, MU, 1)
```

```
Cpl.Sc4(DATA1, 7, MU)
```

```
MU < -11
```

```
DATA2 < -rnorm(100, MU, 0.5)
```

```
Cpl.Sc4(DATA2, 7, MU)
```

```
DATA3 < -rnorm(100, 10, 2)
```

```
Cpl.Sc4(DATA3, 7)
```

---

Cpu	One-sided Process Capability Index (Upper specification only)
-----	---

---

### Description

Calculates the true value of the one-sided Process Capability Index (Upper specification only).

### Usage

```
Cpu(USL, MU, SIGMA)
```

### Arguments

USL	the upper specification limit.
MU	the process mean.
SIGMA	a positive number that indicate the process standard deviation.

**Details** Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . The one-sided Process Capability Index (upper specification only) is defined as:

$$C_{pu} = \frac{USL - \mu}{3\sigma},$$

where  $USL$  is the upper specification limit.

### Value

The function returns the true value of the one-sided Process Capability Index (upper specification only).

## References

Kotz, S., Johnson, N. L. (1993). *Process capability indices*. CRC Press.

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

## See also

Cp

Cpk

Cpl

## Examples

Cpu(13, 10, 1)

Cpu(13, 10, 0.5)

Cpu(13, 10, 2)

---

Cpu.Sc4	Estimator of one-sided Process Capability Index (Upper specification only) based on the sample standard deviation
---------	---

---

## Description

Computes the estimator of one-sided Process Capability Index (upper specification only) using the point estimator based on the sample standard deviation.

## Usage

Cpu.Sc4(Xn, LSL, MU)

## Arguments

Xn	vector of the quality characteristic of interest $X$ ; its length is equal to $n$ .
USL	the upper specification limit.
MU	the process mean. If the process mean is not indicated, the function uses the sample mean, i.e. the mean of Xn.

## Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and a single sample with size  $n$ , the point estimator of the one-side Process Capability Index (upper specification only)

$$C_{pu} = \frac{USL - \mu}{3\sigma},$$

based on the sample standard deviation is defined as:

$$\widehat{C}_{pu.c4} = \frac{USL - \mu}{3\widehat{\sigma}_{c4}},$$

where

$$\widehat{\sigma}_{c4} = \frac{S}{c_4[n]}$$

is the unbiased estimator of the process standard deviation based on the sample standard deviation,

$$S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2},$$

the sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

is the estimator of the process mean,  $x_i$  is the value of the quality characteristic  $X$  for the  $i$ th product, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Finally,  $USL$  is the upper specification limit. Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

It is common to use the sample mean ( $\bar{x}$ ) when the parameter  $\mu$  is unknown, i.e.,  $\hat{\mu} = \bar{x}$ .

### Value

The function returns a value for the one-sided Process Capability Index (upper specification only) point estimator.

### References

Kotz, S., Johnson, N. L. (1993). *Process capability indices*. CRC Press.

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

### See also

Cpu

Cpl.Sc4

### Examples

```
MU < -10
```

```
DATA1 < -rnorm(100, MU, 1)
```

```
Cpu.Sc4(DATA1, 13, MU)
```

```
MU < -11
```

```
DATA2 < -rnorm(100, MU, 0.5)
```



Cpu.Sc4(DATA2, 13, MU)

DATA3 < -rnorm(100, 10, 2)

Cpu.Sc4(DATA3, 13)

---

SIGMA.M.R	Estimator for a process standard deviation based on the sample ranges derived from $m$ samples
-----------	--

---

### Description

Computes the estimator of the process standard deviation using the sample ranges derived from  $m$  samples.

### Usage

SIGMA.M.R(Xn)

### Arguments

Xn matrix (with dimension  $m \times n$ ) of the quality characteristic of interest  $X$ ; each row of Xn corresponds to a sample with length equal to  $n$ .

**Details** Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For  $m$  samples with the same size  $n$ , the point estimator of the process standard deviation based on the sample ranges is defined as

$$\hat{\sigma}_R = \frac{\bar{R}}{d_2[n]}$$

where

$$\bar{R} = \frac{1}{m} \sum_{i=1}^m R_i$$

is the average of the sample ranges defined as

$$R_i = \text{máx}(x_{ij}) - \text{mín}(x_{ij}),$$

$x_{ij}$  denotes the observed value of  $X$  for the  $j$ th product, with  $j = 1, \dots, n$ , in the  $i$ th sample, with  $i = 1, \dots, m$ ,  $d_2[n]$  is defined as

$$d_2[n] = \int_{-\infty}^{\infty} 1 - (Z_y)^n - (1 - Z_y)^n dy.$$

Note that the values of  $d_2[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009).

### Value

The function returns the value of the point estimator of the process standard deviation using the sample ranges derived from  $m$  samples.

### References

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

Woodall W.H., Montgomery D.C. (2000). Using Ranges to Estimate Variability. *Quality Engineering* ; **13(2)**: 211-217.

### See also

SIGMA.M.R.WM

SIGMA.M.Sc4

SIGMA.M.Sp

### Examples

```
DATA1< -matrix(rnorm(100, 0, 1), nc=20)
```

```
SIGMA.M.R(DATA1)
```

```
DATA2< -matrix(rnorm(20, 10, 2), nc=4)
```

```
SIGMA.M.R(DATA2)
```

```
DATA3< -matrix(rnorm(100, 0, 2), nc=5)
```

```
SIGMA.M.R(DATA3)
```

---

SIGMA.M.R.WM	Estimator based on the sample ranges defined by Woodall and Montgomery (2000) of a process standard deviation
--------------	---

---

### Description

Estimates of process standard deviation using the point estimator based on the sample ranges defined by Woodall and Montgomery (2000).

## Usage

SIGMA.M.R.WM(Xn)

## Arguments

Xn matrix (with dimension  $m \times n$ ) of the quality characteristic of interest  $X$ ; each row of Xn corresponds to a sample with length equal to  $n$ .

## Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For  $m$  samples with the same size  $n$ , the point estimator defined by Woodall and Montgomery (2000) of the process standard deviation based on the sample ranges is defined as:

$$\hat{\sigma}_{R.WM} = \frac{d_2[n]}{d_2^*[n]} \bar{R}$$

where

$$\bar{R} = \frac{1}{m} \sum_{i=1}^m R_i$$

is the average of the sample ranges defined as

$$R_i = \text{máx}(x_{ij}) - \text{mín}(x_{ij}),$$

$x_{ij}$  denotes the observed value of  $X$  for the  $j$ th product, with  $j = 1, \dots, n$ , in the  $i$ th sample,  $d_2[n]$  is defined as

$$d_2[n] = \int_{-\infty}^{\infty} 1 - (Z_y)^n - (1 - Z_y)^n dy,$$

$d_2^*[n]$  is defined as

$$d_2^*[n] = \left( d_2^2[n] + \frac{d_3^2[n]}{m} \right)^{1/2},$$

$d_3[n]$  is defined by Woodall and Montgomery (2000). Note that the values of  $d_2[n]$  and  $d_3[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009).

## Value

The function returns the value of the point estimator defined by Woodall and Montgomery (2000) of the process standard deviation using the sample ranges derived from  $m$  samples.

## References

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

Woodall W.H., Montgomery D.C. (2000). Using Ranges to Estimate Variability. *Quality Engineering* ; **13(2)**: 211-217.

## See also

SIGMA.M.R

SIGMA.M.Sc4

SIGMA.M.Sp

## Examples

```
DATA1< -matrix(rnorm(100, 10, 1), nc=20)
```

```
SIGMA.M.R.WM(DATA1)
```

```
DATA2< -matrix(rnorm(20, 10, 2), nc=4)
```

```
SIGMA.M.R.WM(DATA2)
```

```
DATA3< -matrix(rnorm(100, 0, 2), nc=5)
```

```
SIGMA.M.R.WM(DATA3)
```

---

SIGMA.M.Sc4	Estimator of a process standard deviation based on the sample standard deviations derived from $m$ samples
-------------	--

---

## Description

Computes the estimator of the process standard deviation and using the sample standard deviations derived from  $m$  samples.

## Usage

```
SIGMA.M.Sc4(Xn)
```

## Arguments

$X_n$  matrix (with dimension  $m \times n$ ) of the quality characteristic of interest  $X$ ; each row of  $X_n$  corresponds to a sample with length equal to  $n$ .

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For  $m$  samples with the same size  $n$ , the point estimator of the process standard deviation based on the sample standard deviations is defined as:

$$\hat{\sigma}_S = \frac{\bar{S}}{c_4[n]}$$

where

$$\bar{S} = \frac{1}{m} \sum_1^m S_i$$

is the average of the sample standard deviations

$$S_i = \left( \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \right)^{1/2},$$

and where

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$$

is the sample mean for the  $i$ th sample,  $x_{ij}$  denotes the observed value of  $X$  for the  $j$ th product, with  $j = 1, \dots, n$ , in the  $i$ th sample, with  $i = 1, \dots, m$ , and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

### Value

The function returns the value of the point estimator of the process standard deviation using the sample standard deviations derived from  $m$  samples.

### References

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

Chen G. (1997). The mean and standard deviation of the run length distribution of X-bar charts when control limits are estimated. *Statistica Sinica* 7: 789-798.

### See also

SIGMA.M.R

SIGMA.M.R.WM

SIGMA.M.Sp

### Examples

```
DATA1 <- matrix(rnorm(100, 10, 1), nc=20)
```

```
SIGMA.M.Sc4(DATA1)
```

```
DATA2 <- matrix(rnorm(100, 10, 2), nc=4)
```

```
SIGMA.M.Sc4(DATA2)
```

```
DATA3 <- matrix(rnorm(100, 0, 2), nc=5)
```

```
SIGMA.M.Sc4(DATA3)
```

SIGMA.M.Sp	Estimator of a process standard deviation based on the pooled sample standard deviation obtained from $m$ samples
------------	---

### Description

Computes the estimator of the process standard deviation using the pooled sample standard deviation obtained from  $m$  samples.

### Usage

SIGMA.M.Sp(Xn)

### Arguments

Xn                      matrix (with dimension  $m \times n$ ) of the quality characteristic of interest  $X$ ; each row of Xn corresponds to a sample with length equal to  $n$ .

### Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a centered process and  $m$  samples with the same size  $n$ , the point estimator of the process standard deviation based on the sample standard deviations is defined as

$$\hat{\sigma}_{Sp} = \frac{S_p}{c_4[1 + m(n - 1)]},$$

where

$$S_p = \left( \frac{1}{m} \sum_{i=1}^m S_i^2 \right)^{1/2}$$

is the pooled sample standard deviation,

$$S_i = \left( \frac{1}{n - 1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \right)^{1/2},$$

and where

$$\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$$

is the sample mean for the  $i$ th sample,  $x_{ij}$  denotes the observed value of  $X$  for the  $j$ th product, with  $j = 1, \dots, n$ , in the  $i$ th sample, with  $i = 1, \dots, m$ , and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

## Value

The function returns the value of the point estimator of the process standard deviation and using the pooled sample standard deviation obtained from  $m$  samples.

## References

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

Chen G. (1997). The mean and standard deviation of the run length distribution of X-bar charts when control limits are estimated. *Statistica Sinica* **7**: 789-798.

## See also

SIGMA.M.R

SIGMA.M.R.WM

SIGMA.M.Sc4



## Examples

```
DATA1 <- matrix(rnorm(100, 10, 1), nc=20)
```

```
SIGMA.M.Sp(DATA1)
```

```
DATA2 <- matrix(rnorm(100, 10, 2), nc=4)
```

```
SIGMA.M.Sp(DATA2)
```

```
DATA3 <- matrix(rnorm(100, 0, 2), nc=5)
```

```
SIGMA.M.Sp(DATA3)
```

---

SIGMA.Sc4	Estimator of a process standard deviation and based on the sample standard deviation
-----------	--

---

## Description

Computes the estimator of the process standard deviation using the sample standard deviation.

## Usage

```
SIGMA.Sc4(Xn)
```

## Arguments

Xn                      vector of the quality characteristic of interest  $X$ ; its length is equal to  $n$ .

## Details

Let  $X$  be a quality characteristic with mean  $\mu$  and standard deviation  $\sigma$ . For a single sample with size  $n$ , the point estimator of the unbiased estimator of the process standard deviation based on the sample standard deviation is defined as

$$\hat{\sigma}_{c4} = \frac{S}{c_4[n]}$$

where the sample standard deviation is defined as

$$S = \left( \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{1/2},$$

and the sample mean

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

is the estimator of the process mean,  $x_i$  is the value of the quality characteristic  $X$  for the  $i$ th product, and  $c_4[n]$  is defined as

$$c_4[n] = \frac{(\pi/2)^{I_n/2}}{(n-1)^{1/2}} \prod_{i=2}^{n-2} i^{I_{ni}}$$

where  $I_n$  is an indicator variable defined as:

$$I_n = \begin{cases} 1 & \text{if } n \text{ is odd} \\ -1 & \text{if } n \text{ is even} \end{cases}$$

and  $I_{ni}$  takes the following values:

$$I_{ni} = \begin{cases} 1 & \begin{cases} \text{if both } \{n, i\} \text{ are odd} \\ \text{if both } \{n, i\} \text{ are even} \end{cases} \\ -1 & \begin{cases} \text{if } n \text{ is odd and } i \text{ is even} \\ \text{if } n \text{ is even and } i \text{ is odd} \end{cases} \end{cases}$$

Note that the values of  $c_4[n]$ , for various values of  $n$ , are also tabulated (see, for example, Appendix VI from Montgomery, 2009). The customary definition of  $c_4[n]$  can be seen in Chen(1997).

## Value

The function returns the value of the point estimator of the process standard deviation using the sample standard deviation.

## References

Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

Chen G. (1997). The mean and standard deviation of the run length distribution of X-bar charts when control limits are estimated. *Statistica Sinica* **7**: 789-798.

**See also**

Cp.Sc4

Cpk.Sc4

### **Examples**

```
DATA1<-rnorm(100, 10, 1)
```

```
SIGMA.Sc4(DATA1)
```

```
DATA2<-rnorm(100, 10, 2);
```

```
SIGMA.Sc4(DATA2)
```

```
DATA3<-rnorm(100, 21, 2)
```

```
SIGMA.Sc4(DATA3)
```

## Capítulo 5

# Conclusiones, limitaciones y futuras líneas de investigación

El objetivo principal de este capítulo es exponer de forma resumida las principales conclusiones obtenidas en este trabajo, algunas de las limitaciones que presenta, y posibles futuras líneas de investigación relacionadas con el mismo. En primer lugar, se expondrá un resumen de las conclusiones obtenidas en los análisis de los estimadores del *PCI*, y que se derivan de los estudios de simulación. En segundo lugar, se mencionan las conclusiones relacionadas con las aportaciones realizadas al problema de la estimación de dicho índice de capacidad de un proceso. Adicionalmente, se incluye un resumen en el que se muestran los principales resultados relacionados con los objetivos intermedios expuestos en la Sección 1.3. A continuación, se enumeran las principales limitaciones con las que cuenta el presente trabajo. Además, se plantean una serie de futuras líneas de investigación relacionadas con el *PCI* y sobre las que se espera seguir investigando en un futuro.

### 5.1. Conclusiones

El presente trabajo se centra en la estimación de la medida más utilizada por las empresas para aplicar el análisis de capacidad de sus procesos de producción, es decir, el índice de capacidad de un proceso (*PCI*). Este trabajo se puede dividir en dos grandes bloques: análisis de los estimadores tradicionales del índice de capacidad de un proceso en distintos escenarios (Capítulo 2); y aportaciones relacionadas con la estimación de dicho índice de capacidad

(Capítulo 3 y 4). Ambos bloques comparten un objetivo común: mejorar el conocimiento sobre el comportamiento de los estimadores del *PCI* y, de este modo, mejorar la interpretación de la información obtenida a través de este índice cuando debe ser estimado en situaciones reales.

En el primer bloque se ha realizado un exhaustivo análisis del comportamiento de los principales estimadores del *PCI* en determinados escenarios. Debido a la estrecha relación existente entre el *PCI* y la desviación típica del proceso ( $\sigma$ ), se incluye en el análisis los estimadores de la desviación típica en los que se basan los correspondientes estimadores del *PCI*. Las medidas utilizadas en los estudios de simulación correspondientes a este primer bloque para evaluar los estimadores en términos de precisión y eficiencia son, respectivamente, el sesgo relativo (*RB*) y el error cuadrático medio relativo (*RRMSE*). Algunos de los mencionados escenarios tiene como objetivo simular situaciones que ocurren con frecuencia en la realidad y que vulneran alguno o algunos de los supuestos básicos (Supuestos A1, A2 y A3 definidos en la Sección 2.2.1) en los que se basa la estimación del *PCI*. Como se ha argumentado en este trabajo, es común que estos supuestos no puedan cumplirse en la práctica. Por ejemplo, es habitual que las empresas almacenen su producción en lotes finitos de productos. De hecho, esta situación se contempla en otras técnicas incluida en el campo del control estadístico de calidad, como por ejemplo, el muestreo de aceptación. Este escenario descrito implica el incumplimiento del Supuesto A2. Otras situaciones frecuentes en la práctica, también descritas en la Sección 2.2.1, tienen como consecuencia el incumplimiento de los Supuestos A1 y A3. Asimismo, se ha descrito en este trabajo la relación existente entre ambos supuestos. Por tanto, este análisis permite ampliar el conocimiento sobre los principales estimadores del *PCI*. Destacamos que las decisiones sobre un proceso de producción de la empresa se basan, junto con otras, en la información obtenida al aplicar un análisis de la capacidad del proceso. Por tanto, si el *PCI* es estimado, que es lo más común en la práctica, resulta interesante conocer cómo de precisa y eficiente es la estimación que se está realizando.

El análisis realizado en este primer bloque está basado en dos estudios de simulación Monte Carlo. Por un lado, se ha llevado a cabo un estudio empírico en el que se asume la selección de una única muestra para la estimación del *PCI* (Sección 2.2). En este estudio se compara empíricamente los estimadores del *PCI* basados en los estimadores sesgado e insesgado de  $\sigma$  (definidos en la Sección 2.2.2) cuando la muestra es seleccionada de una población tanto finita como infinita. A continuación, se evalúa empíricamente el comportamiento del estimador del *PCI* basado en el estimador sesgado de  $\sigma$  en situaciones donde se satisfacen los supuestos básicos y, por otro lado, situaciones en las que se incumplen uno o más de estos supuestos básicos. Las siguientes con-

clusiones se derivan de los resultados de dicho estudio. En primer lugar, el estimador insesgado de  $\sigma$  y  $PCI$  puede tener un mejor comportamiento que sus homólogos sesgados cuando se asumen muestras de tamaño pequeño. En concreto, en el estudio de simulación realizado dicho estimador presenta mayor precisión en términos de sesgo relativo bajo esta situación. Se ha observado un comportamiento similar en términos de precisión y cobertura empírica de los correspondientes intervalos de confianza. En segundo lugar, se ha comprobado que en procesos no centrados se debe tener en cuenta que los intervalos de confianza para el  $PCI$  pueden obtener coberturas empíricas no deseadas. Por ejemplo, este efecto puede estar presente cuando alguno o algunos de los supuestos no se cumplen. En concreto, en este estudio se ha considerado el incumplimiento del Supuesto A1. Por ejemplo, asumiendo el escenario extremo en el que la característica de calidad sigue una distribución Uniforme en lugar de la requerida distribución Normal, se observan coberturas empíricas en los intervalos de confianza cercanas al 100 %. Sin embargo, si la distribución que se asume es la distribución Gamma (la cual se caracteriza por su asimetría a la derecha), se obtienen coberturas deseables, por lo que el impacto de una distribución con estas características no tiene un efecto importante en los intervalos de confianza para el índice de capacidad de un proceso. Por otro lado, el incumplimiento de los Supuestos A2 y A3 también puede conducir a intervalos de confianza con diferentes coberturas empíricas a las deseadas. Adicionalmente, el incumplimiento del Supuesto A2 puede repercutir en las estimaciones puntuales del índice de capacidad de un proceso, siendo éstas menos precisas. Por ejemplo, en unos de los estudios empíricos realizados se observó que el incumplimiento del Supuesto A2 puede afectar negativamente a los estimadores de este parámetro, puesto que en éstos se obtuvieron sesgos relativos elevados. Sin embargo, estas consecuencias no son relevantes en el caso de poblaciones infinitas o aproximación a poblaciones infinitas (fracciones de muestreo próximas a cero). En resumen, el incumplimiento de Supuestos Básicos en el problema de la estimación del  $PCI$  tiene especial incidencia en el caso de la estimación mediante intervalos de confianza. Aunque disminuyen su rendimiento, en líneas generales los estimadores puntuales del  $PCI$  y de  $\sigma$  tienen un buen comportamiento cuando no se cumple el Supuesto A3. Por tanto, es recomendable interpretar las estimaciones del  $PCI$  con cierta prudencia en procesos que incumplan algún supuesto, considerando el posible efecto que dicho incumplimiento tendrá en la estimación del índice de capacidad.

A modo de extensión, se ha llevado a cabo un segundo estudio de simulación en el que se asume la selección de  $m$  muestras para la estimación del  $PCI$  (Sección 2.3), donde  $m > 1$ . En este caso se compara el comportamiento de cuatro estimadores del  $PCI$  basados en distintos estimadores de  $\sigma$  que son

definidos en la Sección 2.3.1. Al igual que en el primer estudio de simulación, adicionalmente se comparan distintos estimadores de  $\sigma$  en el nuevo contexto de  $m$  muestras. En este estudio se han considerado escenarios en los que la característica de calidad sigue distintas distribuciones: Normal, Gamma y Uniforme. Es decir, se analiza el comportamiento de los estimadores ante el cumplimiento e incumplimiento del Supuesto A1. También se han considerado procesos no centrados. Las principales conclusiones de este estudio se exponen a continuación. El tamaño de las muestras utilizadas en la estimación del *PCI* puede tener influencia sobre la precisión de la estimación. En concreto, los estimadores pueden tener un mejor comportamiento en términos de sesgo empírico cuando los tamaños muestrales ( $n$ ) son superiores a 5. En este sentido, cabe destacar que el comportamiento de los diferentes estimadores mejora a medida que aumenta  $m$  y  $n$ . En general, se observa un peor comportamiento de los estimadores del *PCI* basados en rangos muestrales. Además, dichos estimadores tienen un peor comportamiento en términos de sesgo y eficiencia a medida que aumenta el tamaño de muestra. Por otro lado, se observa que todos los estimadores del *PCI* pueden tener un peor comportamiento en los procesos que incumplen el Supuesto A1. En el caso de datos procedentes de la distribución Gamma, los estimadores pueden tener un comportamiento similar a la situación en que se cumple el Supuesto A1. Por tanto, es pertinente realizar un análisis previo de los datos que nos permita intuir qué impacto tendrá el incumplimiento del Supuesto A1 sobre la estimación del *PCI*.

Si comparamos ambos estudios, se observa que la estimación del *PCI* en el caso de una única muestra mejora notablemente cuando el tamaño muestral es mayor a 10, puesto que se obtienen sesgos relativos aceptables (inferiores al 5%). Sin embargo, en el caso de  $m$  muestras, los sesgos relativos mejoran de manera significativa cuando los tamaños muestrales son superiores a 5 (inferiores a 1%). Recordamos que en el estudio se han considerado los valores  $m = \{10, 20, 30, 50, 100, 1000\}$ . En líneas generales, si atendemos exclusivamente a los tamaños muestrales, la estimación del *PCI* puede mejorar su rendimiento en el caso de  $m$  muestras. Por ejemplo, en los resultados derivados de los estudios de simulación, puede observarse que el *RB* con el estimador del *PCI* basado en la desviación típica muestral en el caso de una única muestra y con tamaño muestral  $n = 5$  es de 18%, mientras que en el caso de  $m = 10$  muestras del mismo tamaño es 1,2%. Sin embargo, puede ser más apropiado comparar los estimadores en función al número de elementos muestrales que aportan información al problema de la estimación del *PCI*. Por ejemplo, si comparamos el caso anterior,  $m = 10$  muestras con un tamaño muestral  $n = 5$ , con el caso de una sola muestra de tamaño  $n = 5 \times 10 = 50$ , se observa un comportamiento similar en la estimación del *PCI* en términos de sesgos

relativos. Se obtienen resultados similares en los estudios de simulación para el escenario de un proceso no centrado. Respecto al incumplimiento del Supuesto A1, si consideramos el uso de la distribución Gamma, los sesgos relativos obtenidos en el caso de  $m$  muestras son similares al caso de una única muestra. Por ejemplo, el sesgo relativo observado en el caso de  $m = 10$  muestras y un tamaño muestral  $n = 5$  es 1,4%, mientras que en el caso de una única muestra de tamaño  $n = 50$  es de 1,76%. Si consideramos el uso de la distribución Uniforme, el estimador basado en una única muestra puede mejorar ligeramente el rendimiento en términos de sesgos relativos. Por ejemplo, si se consideran los mismos valores de  $n$  y  $m$  que en el ejemplo anterior, los sesgos relativos son 0,68% y  $-1,29\%$  para el caso de una y  $m$  muestras, respectivamente. En cualquier caso se observa un mejor comportamiento en términos de *RRMSE* del estimador en cuando se consideran  $m$  muestras.

El segundo bloque del trabajo se compone de aportaciones a la estimación del *PCI*. Por un lado, se han propuesto nuevas expresiones para la función  $c_4[\cdot]$  (Capítulo 3). Esta función es utilizada para estimación de  $\sigma$ , ya que permite obtener un estimador insesgado del parámetro. Como se expone en la Sección 3.2,  $c_4[\cdot]$  depende del tamaño muestral, y sus valores están tabulados o se puede obtener fácilmente a través de la definición de  $c_4[\cdot]$  para valores muestrales pequeños. Sin embargo, siguiendo la comentada definición de  $c_4[\cdot]$ , no es posible calcular su valor para tamaños de muestra elevados, puesto que su expresión que la define depende del cociente de dos funciones Gamma, las cuales tienen a infinito a medida que aumenta el tamaño de muestra. Las nuevas expresiones para la función  $c_4[\cdot]$  propuestas determinan su valor como un producto de fracciones, lo cual hace posible el cálculo de  $c_4[\cdot]$  para tamaños muestrales elevados. En consecuencia, la función  $c_4[\cdot]$  es fácilmente programable en software estadísticos que automaticen el control de calidad en la empresa. Por ejemplo, esta función ha sido programada para llevar a cabo el primer estudio de simulación del Capítulo 2, el cual no hubiese sido posible a partir de las funciones ya implementadas en el programa *R* para el cálculo de la función  $c_4[\cdot]$ .

Por otro lado, se ha elaborado una librería para el software estadístico *R* denominada “Capability”. Esta nueva librería se centra en el análisis de capacidad de un proceso y la estimación del *PCI*, tanto de forma puntual como por intervalos de confianza. Hasta la fecha, el programa estadístico *R* no contaba con una librería que se dedicara de este tema de manera exhaustiva. La librería “Capability” reúne las funciones necesarias para realizar estimaciones del *PCI* usando distintos estimadores y alternativas, y que han sido objeto de estudio en el presente trabajo. Además, esta librería contiene una función, `CONST.C4`, capaz de determinar el valor de  $c_4[\cdot]$  para cualquier tamaño muestral haciendo uso de las nuevas expresiones propuestas.



Para finalizar esta sección, se ha realizado un resumen esquemático de las principales conclusiones derivadas de la consecución de los objetivos descritos en la Sección 1.3:

- O1** Asumiendo una única muestra para realizar las estimaciones de  $\sigma$  y  $PCI$ , el estimador insesgado de  $\sigma$  basado en la desviación típica muestral puede obtener estimaciones más precisas en términos de sesgos relativos que su versión sesgada para tamaños muestrales menores que 10 ( $n < 10$ ). Respecto a estimadores puntuales del  $PCI$ , se obtiene la misma conclusión, es decir, el estimador basado en el estimador insesgado de  $\sigma$  puede obtener sesgos menores para tamaños de muestra pequeños que el estimador basado en el estimador sesgado de  $\sigma$ . En general, ambos estimadores pueden obtener resultados aceptables en la estimación del  $PCI$  a partir de tamaños muestrales iguales o mayores a 15 ( $n \geq 15$ ).
- O2** Los intervalos de confianza del  $PCI$  en procesos no centrados pueden presentar coberturas empíricas mayores a su nivel nominal.
- O3** Los resultados en términos de  $RB$  y  $RRMSE$  derivados del estudio sobre la estimación de  $\sigma$  y del  $PCI$  asumiendo una muestra y el incumplimiento del Supuesto A3 (considerando muestras con valores no incluidos en el intervalo formado por el límite superior e inferior de control), sugieren que este hecho no tiene especial impacto en los estimadores puntuales. Sin embargo, los intervalos de confianza pueden presentar valores de cobertura empírica superiores al nominal.
- O4** En general, el estimador puntual de  $\sigma$  y el estimador puntual de  $PCI$  y por intervalos de confianza pueden tener un peor comportamiento en poblaciones finitas. Los estimadores puntuales pueden presentar sesgos relativos más elevados que en poblaciones infinitas, y los intervalos de confianza coberturas empíricas superiores al nivel nominal.
- O5** Asumiendo fracciones de muestreo próximas a cero, es decir, poblaciones “aproximadamente” infinitas, los estimadores pueden tener un mejor comportamiento que en el caso de poblaciones finitas con fracciones de muestreo elevadas.
- O6** Cuando se cumplen los Supuestos A1 y A2, la distribución empírica del estadístico cuasivarianza muestral se ajusta bien a su distribución teórica,  $\chi^2$  con  $n - 1$  grados de libertad. Por el contrario, en el caso de muestras procedentes de poblaciones finitas, la distribución empírica del estadístico puede no coincidir con su distribución teórica. En el caso de poblaciones

finitas con fracción de muestreo muy pequeñas, la distribución empírica del estadístico se aproxima en gran medida a su distribución teórica.

- O7** El impacto que supone que la característica de calidad no se distribuya según una distribución Normal sobre la estimación del *PCI* puede tener distintas consecuencias en función de la distribución que realmente siga. Por ejemplo, si la distribución de la característica de calidad es una distribución Gamma, el comportamiento de las estimaciones puede ser similar que cuando sigue una distribución Normal. En el caso extremo de datos procedentes de una distribución Uniforme, la cobertura empírica de los intervalos de confianza puede estar cercana al 100 %.
- O8** Los resultados obtenidos en la evaluación del estimador en un entorno real han sido consistentes respecto a los resultados obtenidos en los escenarios simulados. Utilizando el conjunto de datos reales “pistonrings” incluido en la librería “qcc” del software estadístico *R*, se han obtenido estimaciones aceptables para muestreo aleatorio con reemplazamiento y las observaciones estaban en situación de bajo control. Este esquema de muestreo permite garantizar la independencia entre elementos muestrales. El hecho de que las muestras sean seleccionadas sin reemplazamiento (dependencia entre elementos muestrales) o que se consideren observaciones en situación de fuera de control, puede repercutir especialmente en la calidad de los intervalos de confianza, obteniéndose coberturas distintas a las deseadas.
- O9** Asumiendo  $m$  muestras para realizar las estimaciones de  $\sigma$  y *PCI*, los estimadores estudiados pueden presentar sesgos relativos elevados con tamaños muestrales menores a 5. Los estimadores basados en los rangos muestrales pueden tener un peor comportamiento que los basados en la desviación típica muestral. En particular pueden ser menos precisos. Los estimadores pueden presentar un comportamiento similar al descrito en procesos no centrados.
- O10** El incumplimiento de la hipótesis de normalidad en los datos puede afectar de forma diferente según la distribución real de los mismos. Por ejemplo, si los datos siguen una distribución Gamma, puede no tener un elevado impacto sobre el rendimiento de los estimadores. Si por el contrario, nos encontramos en el caso extremo de datos procedentes de una distribución Uniforme, los estimadores basados en rangos pueden tener un mal comportamiento y, además, puede empeorar a medida que aumenta el tamaño muestral.

- O11** Se han propuesto dos nuevas expresiones para el cálculo de la función  $c_4[\cdot]$ . La primera expresión permite obtener  $c_4[\cdot]$  para un determinado tamaño muestral a partir del valor de la propia función evaluada respecto a un tamaño muestral inferior. La segunda expresión calcula el valor de  $c_4[\cdot]$  sin necesidad de cálculos previos. A partir de ambas expresiones es posible calcular el valor de  $c_4[\cdot]$  para cualquier tamaño muestral y, como consecuencia, es posible programar una función que calcule el valor exacto de  $c_4[\cdot]$  en software estadísticos.
- O12** Se ha desarrollado una librería, denominada “Capability”, que se centra de forma exhaustiva en la estimación del *PCI*. Esta librería se ha desarrollado para el entorno del software estadístico *R*. Se espera que dicha librería pueda ser usada por las empresas a modo de herramienta con la que poder efectuar estimaciones del *PCI*, y de esta forma contribuir en la fase del control estadístico de la calidad de su producción empresarial.

## 5.2. Limitaciones y Futuras líneas de investigación

En esta sección se exponen las limitaciones y futuras líneas de investigación. En primer lugar, se mencionan algunas limitaciones con las que cuenta el presente trabajo y las consecuentes líneas de investigación que se derivan de ellas. Posteriormente, se enumeran algunas líneas de investigación relacionadas con nuevas metodologías para la estimación del *PCI*.

En primer lugar, este trabajo asume, por simplicidad, que el control de calidad es aplicado a un producto. Sin embargo, el conjunto de técnicas cuantitativas que conforman este campo es susceptible de aplicación a otros ámbitos como el sector servicios. Por ejemplo, en un servicio de reparto a domicilio podría identificarse la característica de calidad con el tiempo de espera hasta recibir el pedido. Otro ejemplo puede ser el propuesto por Montgomery (2009, p.306), donde la característica de calidad indica el número real de órdenes de compras emitidas por semana. Por tanto, puede ser interesante analizar el comportamiento de los diferentes índices de capacidad cuando se aplica a procesos del sector servicios, y contar para estas investigaciones con bases de datos reales.

En segundo lugar, se ha considerado una sola característica de calidad de interés para el estudio de la capacidad del proceso. En determinadas ocasiones,

existe más de una característica de interés involucrada en el proceso. Montgomery (2009, p.495) propone un rodamiento como ejemplo de artículo con más de una característica de calidad, el cual tendrá utilidad dependiendo de la longitud de su diámetro interior y su diámetro exterior. En otras palabras, dicho producto tiene dos características de calidad que deben ser controladas, la longitud del diámetro interior y del diámetro exterior. En esta situación, se recurre a los índices de capacidad multivariantes. Estos índices han sido estudiados por numerosos autores como Chen (1994), Hubele et al. (1991), Pan y Lee (2010), Rencher (2002), Scagliarini (2011), Shahriari et al. (1995), Taam et al. (1993), Wang (2005), Wang y Chen (1998), Wang et al. (2000), etc. En definitiva, como futura línea de investigación se propone el estudio de los índices de capacidad multivariantes basados en varias características de calidad.

En tercer lugar, este trabajo de investigación se ha centrado en los índices de capacidad de un proceso  $C_p$  y  $C_{pk}$ . En este sentido, ha de tenerse en cuenta la existencia de otros índices que evalúan la aptitud del proceso y que no han sido analizados en este trabajo. Por ejemplo, los índices de capacidad unilaterales han sido definidos en el Capítulo 1 y se caracterizan por considerar un único límite de especificación, ya sea superior o inferior. Por otro lado, existen algunas alternativas de índices para procesos no centrados como el  $C_{pm}$ , analizado por Chan et al. (1988) y Boyles (1991) o el índice  $C_{pkm}$ , propuesto por Pearn et al. (1992).

Bajo unos supuestos básicos, en el Capítulo 2 se han definido los estimadores puntuales del  $PCI$  y por intervalos de confianza. Entre los mencionados los supuestos básicos, el Supuesto A1 hace referencia a que la característica de calidad se distribuye según una distribución Normal. Posteriormente, se ha analizado el comportamiento de dichos estimadores ante la falta de normalidad en los datos. Puede ser interesante considerar algunos índices de capacidad robustos. Por ejemplo, Luceño (1996) propone el índice de capacidad  $C_{pc}$  para procesos en los que la característica de calidad no tiene distribución Normal. Siguiendo esta línea, una futura línea de investigación puede ser definir estimadores puntuales y por intervalos de confianza de índices de capacidad aplicables a procesos en los que la variable asociada a la característica de calidad siguen una distribución particular distinta a la Normal, por ejemplo, si se distribuye según una ley de probabilidad Gamma.

En el Capítulo 4 se ha presentado la librería para R “Capability”. Aunque esta librería está completa en el sentido de incluir los índices de capacidad habituales y la estimación de éstos de forma puntual y mediante intervalos de confianza, sería interesante completarla con otros índices de capacidad más complejos, unilaterales, etc. Es decir, se espera completar y mejorar la librería

antes de proponerla para su inclusión en la lista de librerías ofertadas en el programa *R*. Por ejemplo, la idea es incluir en dicha librería bases de datos reales, tal como “pistonrings”, la cual si está incluida en la librería “qcc”. Del mismo modo, se tiene intención de incluir funciones relacionadas con las líneas de investigación descritas en esta Sección, como por ejemplo incluir índices de capacidad no considerados en el trabajo como  $C_{pm}$  o  $C_{pc}$ , estimación del *PCI* mediante nuevas metodologías, estimación por intervalos confianza del *PCI* para datos que no procedan de una distribución normal, etc.

A continuación se enumeran algunas líneas de investigación relacionadas con la estimación del *PCI*, y que se esperan que puedan ser analizadas en un futuro próximo.

Como se ha mencionado a lo largo del presente trabajo, las estimaciones del *PCI* son obtenidas mediante información muestral. Se contemplan, tanto en este trabajo como en la literatura sobre el tema, dos casos: seleccionar una única muestra o seleccionar  $m$  muestras. Sin embargo, en ambos casos se asume el esquema de muestreo aleatorio simple sin reemplazamiento. Es bien conocido que el esquema de muestreo utilizado para la selección de individuos o artículos puede producir resultados satisfactorios, es decir, las estimaciones podrían ser más precisas y tener sesgos más pequeños. Por ejemplo, es conocido que, bajo determinadas condiciones, el muestreo estratificado o el muestreo basado en información auxiliar son más precisos que el tradicional muestreo aleatorio simple. Algunos textos que tratan el muestreo estratificado y el muestreo con información auxiliar son Singh y Mangat (2013) y Särndal et al. (1992, p.229). Además, podría analizarse la posibilidad aplicar otros diseños muestrales como por ejemplo el muestreo sistemático, por conglomerados o bifásico. En este sentido, una futura línea de investigación, en la que estamos trabajando actualmente, es la estimación del *PCI* mediante el uso de distintas técnicas para la selección de muestras.

Otra línea de investigación en la que se espera trabajar en el futuro es la propuesta de nuevos estimadores para el *PCI* con mejores propiedades. Para esto, se propone definir estimadores para el *PCI* haciendo uso de uso de nuevas técnicas tal como el método de verosimilitud pseudo empírica (Chen y Sitter 1999), el método de calibración (Deville y Särndall 1992, Harms y Duchesne 2006), etc. En este sentido, pero enfocándonos en la estimación de la varianza y la construcción de intervalos de confianza, también puede resultar muy interesante investigar la estimación de la varianza de estimadores del *PCI* mediante métodos de remuestreo comunes tal como el Jackknife (Shao y Wu 1989, Särndal et al. 1992, Dalgleish 1995, Shao y Tu 1995, Wolter 2007) o el Bootstrap (Bickel y Freedman 1984, Stine 1989, Efron y Tibshirani 1993,

Kovačevik et al. 1995, Biewen 2002, Davidson y Flachair 2007), o bien, otras técnicas también muy eficaces como el método BRR (Shao y Wu 1992), la técnica de linealización (Berger y Skinner 2003, Osier 2009), el ratio de verosimilitud empírica (Berger y De La Riva Torres 2012), el bootstrap reescalado (Kovar et al. 1988, Rao y Wu 1988, Rao et al. 1992), etc. Algunas de estas técnicas han sido aplicadas en el problema de la estimación de varianzas de estimadores de parámetros que se definen como un ratio, es decir, como es el caso del *PCI*, por lo que se esperan resultados satisfactorios en este sentido.

Para finalizar, también podría tener una relevante aplicabilidad e interés el tratamiento de información faltante en el problema de la estimación del índice de capacidad. A modo de ejemplo, a lo largo de este trabajo se ha comprobado que numerosas técnicas, tal como los gráficos de control o la estimación del *PCI* basada en  $m$  muestras, están basadas en varias muestras con un mismo tamaño. Aunque estas técnicas se pueden aplicar a muestras con distintos tamaños muestrales, el hecho de disponer de muestras con el mismo tamaño facilita la interpretación de resultados y suele generar conclusiones más fiables. Aunque una empresa o un determinado estudio esté basado en muestras con el mismo tamaño, existen numerosas razones por las que se puede perder información (véase Rubin 1996), y en consecuencia, las muestras pueden llegar a tener distintos tamaños. Por ejemplo, se puede producir una pérdida accidental de información, estropearse o romperse el aparato de medida, falta de personal en un determinado momento, etc. En esta situación, puede resultar interesante el uso de alguna técnica de imputación que sustituya el dato faltante por un valor aproximado, y para ello se debe hacer un buen uso del resto de información disponible en las muestras. Las técnicas de imputación se pueden clasificar en dos grupos: imputación simple (Sedransk, 1985, Kalton y Kasprzyk, 1986 y Little y Rubin, 2002) e imputación múltiple (Rubin 1978). La primera tiene una implantación más sencilla en comparación con las técnicas de imputación múltiple. Como consecuencia, la imputación simple ha sido una de las herramientas más conocidas y aceptadas para el tratamiento de la falta de respuesta. Las técnicas simples de imputación se pueden dividir en: aleatorias y determinísticas. Éstas últimas generalmente subestiman más las varianzas que las aleatorias, pero proporcionan estadísticos más precisos. Como métodos de imputación determinísticos destacamos el método NNI (Chen y Shao 2000), el de razón (Rao 1996) o el de regresión (Healy y Westmacott 1956), mientras que el método RHD (Random Hot Deck) es uno de los métodos probabilísticos más populares. Por último, destacamos la importancia, en el contexto de los métodos de imputación, de conocer los riesgos derivados de un mecanismo de respuesta no uniforme y de su correcto tratamiento (véase Rancourt et al. 1994 y Hu et al. 2001). La diversidad de técnicas, la comple-

alidad del problema del tratamiento de datos faltantes, y los beneficios que se pueden obtener con el uso de estas herramientas, hacen que la aplicación de este tema en el problema de la estimación del índice de capacidad sea una línea de investigación preferente en un futuro próximo.

# Bibliografía

- [1] Anis M.Z. (2008) Basic Process Capability Indices: An Expository Review. *International Statistical Review*; **76**: 347-367.
- [2] Aslam M., Azam M., Wu C., Jun C. (2015) Mixed acceptance sampling plans for product inspection using process capability index. *Quality Control and Applied Statistics*; **60**: 455-486.
- [3] Beirlant J., Goegebeur Y., Segers J., Teugels J. (2006). *Statistics of extremes: theory and applications*. John Wiley & Sons.
- [4] Berger Y.G., Skinner C.J. (2003). Variance estimation of a low-income proportion. *Journal of the Royal Statistical Society Series C*; **52**: 457-468.
- [5] Berger Y. G., De La Riva Torres O. (2006). *A unified theory of empirical likelihood ratio confidence intervals for survey data with unequal probabilities*. Proceedings of the Survey Research Method Section of the American Statistical Association, Joint Statistical Meeting, San Diego, 15pp.
- [6] Besterfield D. (2009). *Control de Calidad*(9th edn) Prentice Hall. Hispanoamericana S.A.
- [7] Besterfield D. (2014). *Quality Improvement*(9th edn) Harlow : Pearson education.
- [8] Besseris G. (2014). Robust process capability performance: An interpretation of key indices from a nonparametric viewpoint. *The TQM Journal*; **26 (5)**: 445-462.
- [9] Bickel P.J., Freedman D.A. (1984). Asymptotic normality and the bootstrap in stratified sampling. *The Annals of Statistics*; **12**: 470-482.
- [10] Biewen M. (2002). Bootstrap inference for inequality, mobility and poverty management. *Journal of Econometrics*; **108**: 317-342.



- [11] Bissell A.F. (1990). How Reliable Is Your Capability Index? *Applied Statistics*; **39** (3): 331-340.
- [12] Boyles R.A. (1991). The Taguchi Capability Index *Journal of Quality Technology*; **23** (2): 107-126.
- [13] Breygogle F.W. (2003). *Implementing Six Sigma: Smarter Solutions Using Statistical Methods*. Wiley.
- [14] Chakraborti S., Human S.W., Graham M.A. (2008). Phase I Statistical Process Control Charts: An Overview and Some Results. *Quality Engineering*; **21** (1): 52-62.
- [15] Chan L.K., Cheng S.W., Spiring F.A. (1988). A New Measure of Process Capability: Cpm. *Journal of Quality Technology*; **20** (3): 160-175.
- [16] Chan L.K., Cheng S.W., Spiring F.A. (1988). A graphical technique for process capability. *ASQC Quality Congress Transactions*; **42**: 268-275.
- [17] Chen G. (1997). The mean and standard deviation of the run length distribution of X-bar charts when control limits are estimated. *Statistica Sinica*; **7**: 789-798.
- [18] Chen J.P., Ding C.G. (2001). A new process capability index for non-normal distributions. *International Journal of Quality and Reliability Management*; **18**: 762-770.
- [19] Chen J., Shao J. (2000). Nearest neighbor imputation for survey data. *Journal of Statistics*; **16**: 113-131.
- [20] Chen J., Sitter R.R. (1999). A pseudo empirical likelihood approach to the effective use of auxiliary information in complex surveys. *Statistica Sinica*; **9**: 385-406.
- [21] Chen K.S., Huang M.L., Li R.K. (2001). Process capability analysis for an entire product. *International Journal of Production Research*; **39** (17): 4077-4087.
- [22] Chen K.S., Pearn W.L., Lin P.C. (2003). Capability measures for processes with multiple characteristics. *Quality and Reliability Engineering International*; **19**: 101-110.
- [23] Chou Y.M., Owen D.B. (1989). On the distribution of the estimated process capability indices. *Communications in Statistics - Theory and Methods*; **18**: 4549-4560.

- [24] Chou Y.M., Owen D.B., Borrego A.S.A. (1990). Lower confidence limits on process capability indices. *Journal of Quality Technology*; **22**: 223-229.
- [25] Cochran W.G. (1977). *Sampling Techniques*. New York: John Wiley.
- [26] Dalglish L.I.(1995). Software review: Bootstrapping and jackknifing with BOJA. *Statistics and Computing*; **5**: 165-174.
- [27] Davidson R., Flachaire E.(2007). Asymptotic and bootstrap inference for inequality and poverty measures. *Journal of Econometrics*; **141**: 141-166.
- [28] Deville J.C., Särndal C.E.(1992). Calibration estimators in survey sampling. *Journal of the American Statistical Association*; **87**: 376-382.
- [29] Dodge, H.F., Romig, H.G. (1929). A method of sampling inspection. *Bell Labs Technical Journal*; **8 (4)**: 613-631.
- [30] Dovich R.A. (1992). *Quality engineering statistics*. Milwaukee, Wisconsin: ASQ Quality Press.
- [31] Duncan A.J. (1986). *Quality control and industrial statistics*(5th edn) Homewood, IL: Richard D. Irwin.
- [32] Efron B., Tibshirani R.J. (1993). *An introduction to the Bootstrap*. Chapman & Hall, London.
- [33] English J.R., Taylor G.D.(1993). Process Capability Analysis: A Robustness Study. *International Journal of Production Research*; **31 (7)**: 1621-1635.
- [34] Evans J.R., Lindsay W.M. (1999). *The management and control of quality*.(8th edn) South Western College.
- [35] Franklin J., Daoud A. (2011). *Proof in Mathematics: An Introduction*. Sydney: Kew Books.
- [36] Franklin L.A., Wasserman G.S. (1992). A note on the conservative nature of the tables of lower confidence limits for  $C_{pk}$  with a suggested correction. *Statistics-Simulation and Computation*; **21 (4)**: 1165-1169.
- [37] Guirguis G.H., Rodríguez R.N. (1992). Computation of Owen's Q function applied to process capability analysis. *Journal of Quality Technology*; **24 (4)**: 236-246.
- [38] Gujarati D. N. (2009). *Basic econometrics*. McGraw-Hill Education.

- [39] Hann G.J., Hill W.L., Hoerl R.W., Zinkgraf S.A. (1999). The Impact of Six Sigma Improvement-A Glimpse in the Future of Statistics. *The American Statistician*; **53** (3): 208-215.
- [40] Harms T., Duchesne P. (2006). On calibration estimation for quantiles. *Survey Methodology*; **32**: 37-52.
- [41] Healy, M.J.R. y Westmacott, M. (1956). Missing values in experiments analysed on automatic computers. *Applied Statistics*; **5**: 203-206.
- [42] Heavlin W. D. (1988). Statistical properties of capability indices. *Technical report*; **320**, Tech. Library, Advanced Micro Devices, Inc., Sunnyvale, California.
- [43] Hermes H. (1973). *Introduction to Mathematical Logic*. Hochschultext. London: Springer.
- [44] Herrerías R., Palacios F., Callejón J. (2012). *Técnicas Cuantitativas para la Inferencia*. Delta ediciones.
- [45] Hu M., Salvucci S., Lee, R. (2001). *A Study of Imputation Algorithms*. Working Paper No. 200117. Washington DC: U.S. Department of Education, National Center for Education Statistics, 2001. 27 Stata Statistical Software.
- [46] Hurwitz A., Mathur M. (1992). A very simple set of process control rules. *Quality Engineering*; **5** (1): 21-29.
- [47] Jensen W.A., Jones-Farmer L.A., Champ C.W., Woodall W.H. (2006). Effects of parameter estimation on control chart properties: a literature review. *Journal of Quality Technology*; **38** (4): 349-364.
- [48] Jones L.A., Champ C.W., Rigdon S.E. (2001). The performance of exponentially weighted moving average charts with estimated parameters. *Technometrics*; **43**: 156-167.
- [49] Kane V.E. (1986). Process Capability Indices. *Journal of Quality Technology*; **18** (1): 41-52.
- [50] Kalton, G. y Kasprzy, D. (1986). The treatment of missing survey data. *Survey Methodology*; **12**: 1-16.
- [51] Kotz S., Johnson N.L. (1993). *Process capability indices*. CRC Press.
- [52] Kotz S., Johnson N.L. (2002). Process Capability Indices: A Review 1992-2000. *Journal of Quality Technology*; **34** (1): 2-19.

- [53] Kotz S., Lovelace C.R. (1998). *Process Capability Indices in Theory and Practice* Arnold, London.
- [54] Kovar J.G., Rao J.N.K., Wu C.F.J. (1992). Bootstrap and other methods to measure errors in survey estimates. *The Canadian Journal of Statistics*; **16**: 25-45.
- [55] Kovačevik M.S., Yung W., Pandher G.S. (1995). *Estimating the sampling variances of measures of income inequality and polarization - An empirical study*. Statistic Canada, Methodology Branch Working Paper, HSMD-95-007E.
- [56] Kushler R.H., Hurley P. (1992). Confidence Bounds for Capability Indices. *Journal of Quality Technology*; **24** (4): 188-195.
- [57] Li H., Owen D.B., Borrego A.S.A. (1990). Lower confidence limits on process capability indices based on the range. *Communications in Statistics - Simulation and Computation*; **19**: 1-24.
- [58] Linderman K., Schroeder R.G., Zaheer S., Choo A.S. (2003). Six Sigma: a goal-theoretic perspective. *Journal of Operations management*; **21** (2): 193-203.
- [59] Little R.J.A., Rubin D.B. (2002). *Statistical analysis with missing data*. (2nd edn). New York: John Wiley & Sons, Inc.
- [60] Luceño A. (1996). A Process Capability Ratio with Reliable Confidence Intervals. *Communications in Statistics - Simulation and Computation*; **25** (1): 235-246.
- [61] Lucier G.T., Seshadri S. (2001). GE Takes Six Sigma Beyond the Bottom Line. *Strategic Finance*; **82** (11): 40-46.
- [62] Luko S.N. (1996). Concerning the estimators  $\bar{R}/d_2$  and  $\bar{R}/d_2^*$  in estimating variability in a Normal Universe. *Quality Engineering*; **8** (3): 481-487.
- [63] Mitra A. (2008). *Fundamentals of quality control and improvement* Hoboken, N.J. Wiley.
- [64] Montgomery D.C. (1985). *Introduction to statistical quality control*. New York, Wiley.
- [65] Montgomery D.C. (2009). *Statistical Quality Control. A modern introduction*. (6th edn). New York, Wiley.

- [66] Muñoz J.F., Álvarez-Verdejo E., Rueda M.M. (2014). Optimum design-based ratio estimators of the distribution function. *Journal of Applied Statistics*; **41** (7): 1395–1407.
- [67] Muñoz J.F., Álvarez-Verdejo E., Pérez-Aróstegui M.N., Gutiérrez-Gutiérrez L. (2016). Empirical comparisons of X-bar charts when control limits are estimated. *Quality and Reliability Engineering International*; **32**: 453-464.
- [68] Nagata Y., Nagahata H. (1994). Approximation formulas for the lower confidence limits of process capability indices. *Okayama Economic Review*; **25** (4): 301-314.
- [69] Nezhad M., Niaki S. (2013). A new acceptance sampling policy based on number of successive conforming items. *Communications in Statistics - Theory and Methods*; **42**: 1542-1552.
- [70] Osier, G. (2009). Variance estimation for complex indicators of poverty and inequality using linearization techniques. *Journal of the European Survey Research Association*; **3**: 167-195.
- [71] Ott E.R. (1975). *Process quality control*. New York: McGraw-Hill.
- [72] Pandee P.S., Neuman R.P., Cavanagh R.R. (2003). *Las claves de Seis Sigma: la implantación con éxito de una cultura que revoluciona el mundo empresarial*. Madrid: McGraw Hill.
- [73] Pearn W.L., Kotz S., Johnson N.L. (1992). Distributional and Inferential Properties of Process Capability Indices. *Journal of Quality Technology*; **24** (4): 216-231.
- [74] Porter L.J., Oakland J.S. (1991). Process capability indices-an overview of theory and practice. *Quality and Reliability Engineering International*; **7**: 437-448.
- [75] R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- [76] Rancourt E., Lee H., Särndal, C.E. (1994). Bias correction for survey estimates from data with ratio imputed values for confounded nonresponse. *Survey Methodology*; **20**: 137-147.
- [77] Rao, J.N.K. (1996). On variance estimation with imputed survey data (with discussion). *Journal of the American Statistical Association*; **91**: 499-520.

- [78] Rao J.N.K., Kovar J.G., Mantel H.J. (1990). On estimating distribution function and quantiles from survey data using auxiliary information. *Biometrika*; **77**: 365-375.
- [79] Rao G.S., Rosaiah K., Babu M.S. Sivakumar, D. C. (2016). New Acceptance Sampling Plans Based on Percentiles for Exponentiated Fréchet Distribution. *Economic Quality Control*; **31 (1)**: 37-44.
- [80] Rao J.N.K., Wu C.F.J. (1988). Resampling inference with complex survey data. *Journal of the American Statistical Association*; **83**: 231-241.
- [81] Rao J.N.K., Wu C.F.J., Yue K. (1992). Some recent work on resampling methods for complex surveys. *Survey Methodology*; **18**: 209-217.
- [82] Recchia D.R., Barbosa E.P., Goncalves E.J. (2010). IQCC: Improved Quality Control Charts. Extraído el 30 de mayo de 2017, de <http://CRAN.R-project.org/package=IQCC>.
- [83] Rodríguez R.N. (1992). Recent Developments in Process Capability Analysis. *Journal of Quality Technology*; **24 (4)**: 176-187.
- [84] Rubin, D.B. (1978). *Multiple imputations in sample surveys. A phenomenological Bayesian approach to nonresponse*. Proceedings of the survey Research Methods Section, American Statistical Association. 20-34.
- [85] Rubin, D.B. (1996). Multiple imputation after 18+ years. *Journal of the American Statistical Association*; **91**: 473-489.
- [86] Rueda M.M., Muñoz J.F. (2011). Estimation of poverty measures with auxiliary information in sample surveys. *Quality & Quantity*; **45**: 687-700.
- [87] Särndal C.E., Swensson B., Wretman J.H. (1992) *Model Assisted Survey Sampling*. Springer-Verlag, New York.
- [88] Schilling E.G., Nelson P.R. (1976). The effect of Nonnormality on the Control Limits of  $\bar{x}$  Charts. *Journal of Quality Technology*; **4**: 183-188.
- [89] Scrucca L. (2004). qcc: an R package for quality control charting and statistical process control. *R News* ; **4 (1)**: 11-17.
- [90] Sedransk, J. (1985). *The objective and practice of imputation*. First Annual Res. Conf., Washington, D.C.: Bureau of the Census. 445-452.
- [91] Shao J., Tu D. (1995) *The Jackknife and Bootstrap*. Springer-Verlag, New York.

- [92] Shao J., Wu C.F.J. (1989). A general theory for jackknife variance estimation. *The Annals of Statistics*; **17**: 1176-1197.
- [93] Shao J., Wu C.F.J. (1992). Asymptotic properties of the balanced repeated replication method for sample quantiles. *The Annals of Statistics*; **20**: 1571-1593.
- [94] Shewhart W.A. (1931) *Economic control of quality of manufactured product*. Milwaukee, ASQC Quality Press.
- [95] Silva P.L.D., Skinner C.J. (1995). Estimating distribution function with auxiliary information using poststratification. *Journal of Official Statistics*; **11**: 227-294.
- [96] Somerville S.E., Montgomery D.C. (1996). Process Capability Indices and Nonnormal Distributions. *Quality Engineering*; **9 (2)**: 305-316.
- [97] Spiring F., Leung B., Cheng S., Yeung A. (2003). A Bibliography of Process Capability Papers. *Quality and Reliability Engineering International*; **19 (5)**: 445-460.
- [98] Stine R.(1989). An introduction to bootstrap methods: examples and ideas. *Sociological Methods and Research*; **18**: 243-291.
- [99] Vardeman S.B. (1999). A brief tutorial on the estimation of the process standard deviation. *IIE Transact*; **31**: 503-507.
- [100] Wheeler D.J. (1995) *Advanced topics in statistical process control*. Knoxville, TN: SPC press.
- [101] Wilk M.B., Gnanadesikan R (1968). Probability plotting methods for the analysis for the analysis of data. *Biometrika*; **55 (1)**: 1-17.
- [102] Wolter K.M. (2007). *Introduction to variance estimation*.(2nd edn) Springer.
- [103] Woodall W.H., Montgomery D.C. (2000). Using Ranges to Estimate Variability. *Quality Engineering*; **13 (2)**: 211-217.
- [104] Wooldridge J. M. (2015). *Introductory econometrics: A modern approach*. Nelson Education.
- [105] Yeh A.B., Bhattacharya S. (1998). A robust process capability index. *Communications in Statistics - Simulation and Computation*; **27 (2)**: 565-589.

- [106] Yourstone S., Zimmer W. (1992). Non-normality and the Design of Control Charts for Averages. *Decision Sciences*; **23**: 1099-1103.