

# Aplicación de la metodología no paramétrica Bootstrap en el control de calidad del proceso de envasado del café

## Application of non-parametric Bootstrap methodology in quality control of the coffee packaging process

Recibido: junio 20 de 2020 | Revisado: julio 20 de 2020 | Aceptado: agosto 17 de 2020

CELSO GONZALES CHAVESTA<sup>1</sup>  
VERÓNICA PACHECO OTÁROLA<sup>2</sup>  
MANUEL BALLENA<sup>3</sup>

### RESUMEN

La investigación tiene como objetivo mostrar la eficiencia de la metodología Bootstrap en la construcción de límites de los diagramas de control del promedio ( $\bar{X}$ ) - Rango (R), su empleo en las gráficas de control de Wu y Wang (1996) y comparar los resultados de esta metodología con la forma clásica. Se encontró que los límites de control para el gráfico R con el intervalo normal Bootstrap para 1000 (LCL = 0; UCL=13.39) y 10000 (LCL = 0; UCL=13.49), presenta un ligero incremento en la variación del peso; mientras que, con el intervalo de percentil Bootstrap para 1000 (LCL = 0.1; UCL=12.1) y 10000 (LCL = 0.1; UCL=11.8), presenta una ligera disminución en la variabilidad de los pesos respectivamente. Así mismo, los límites de control del promedio Bootstrap con el intervalo normal Bootstrap para 1000 (LCL = 195.54; UCL=207.42) y 10000 (LCL = 195.8; UCL=207.2) el cual presenta una ligera estabilidad en el promedio de los pesos mientras que, con el intervalo de percentil Bootstrap para 1000 (LCL = 196.6; UCL=206.7) y 10000 (LCL = 196.6; UCL=206.3). Por tanto, el diagrama de control promedio-rango, cuando el supuesto de normalidad no se cumple para muestras pequeñas ha mostrado ineficacia en la detección e identificación de causas especiales en un proceso. La investigación presenta una metodología no paramétrica que identifica la señal fuera de control (cumpla o no el supuesto de normalidad), a través de la metodología Bootstrap en la construcción de límites de control promedio-rango.

**Palabras clave:** Bootstrap, gráficos de control, envasado de café

### ABSTRACT

The objective of the research is to show the efficiency of the Bootstrap methodology in the construction of limits of the control diagrams of the mean ( $\bar{X}$ ) - Range (R), its use in the control charts of Wu and Wang (1996) and to compare the results of this methodology with the classical form. It was found that the control limits for the R chart with the Bootstrap normal interval for 1000 (LCL = 0; UCL = 13.39) and 10000 (LCL = 0; UCL = 13.49), present a slight increase in the weight variation; while, with the Bootstrap percentile interval for 1000 (LCL = 0.1; UCL = 12.1) and 10000 (LCL = 0.1; UCL = 11.8), it presents a slight decrease in the variability of the weights respectively. Likewise, the control limits of the Bootstrap average with the Bootstrap normal interval for 1000 (LCL = 195.54; UCL = 207.42) and 10000 (LCL = 195.8; UCL = 207.2) which presents a slight stability in the average of the weights while than, with the Bootstrap percentile interval for 1000 (LCL = 196.6; UCL = 206.7) and 10000 (LCL = 196.6; UCL = 206.3). Therefore, the mean-range control diagram, when the assumption of normality is not fulfilled for small samples, has shown ineffectiveness in the detection and identification of special causes in a process. The research presents a non-parametric methodology that identifies the out-of-control signal (whether or not the assumption of normality), through the Bootstrap methodology in the construction of average-range control limits.

**Key words:** Bootstrap, control charts, coffee packaging

- 1 Universidad Agraria La Molina  
cgonzales@lamolina.edu.pe
- 2 Universidad Agraria La Molina
- 3 Universidad de San Martín de Porres

© Los autores. Este artículo es publicado por la Revista Campus de la Facultad de Ingeniería y Arquitectura de la Universidad de San Martín de Porres. Este artículo se distribuye en los términos de la Licencia Creative Commons Atribución No-comercial – Compartir-Igual 4.0 Internacional (<https://creativecommons.org/licenses/by/4.0/>), que permite el uso no comercial, distribución y reproducción en cualquier medio siempre que la obra original sea debidamente citada. Para uso comercial contactar a: [revistacampus@usmp.pe](mailto:revistacampus@usmp.pe).

<https://doi.org/10.24265/campus.2020.v25n30.11>

## Introducción

El Bootstrap es un método de remuestreo propuesto por Bradley Efron en 1979 que ofrece una aproximación para la inferencia y el contraste de hipótesis. Mediante el uso del remuestreo se estima la distribución empírica del estadístico y a partir de ella se realiza la inferencia sobre el parámetro. Puede considerarse como un tipo especial de simulación basada en los datos, esto es, simulación a partir de una estimación de la población en el que la muestra se trata como si representara exactamente toda la población.

En dicha técnica se puede apreciar como los procesos computacionales sustituyen el análisis teórico. Wu y Wang (1996) exponen el uso de la metodología Bootstrap en los gráficos de control  $\bar{X}$  y R para muestras de tamaño “n” menor a cuatro como una alternativa a los diagramas clásicos de Shewhart los cuales se basan en la asunción de normalidad de la distribución y generalmente para “K” muestras ( $K \geq 25$ ) de tamaño  $n = 5$ . La metodología Bootstrap permite a los usuarios especificar y ajustar el error Tipo I, de tal forma que el poder de detección de los diagramas puede ser mejorado mientras la tasa de falsa alarma se mantiene a un nivel manejable.

Por otro lado, cualquier tipo de carta de control debe de cumplir al menos las siguientes cuatro condiciones: 1) Una sola respuesta a la pregunta ¿está el proceso bajo control?, 2) Especificación del error tipo I global, 3) Se deben tomar en cuenta la relación entre las características a controlar, y 4) Se debe dar un procedimiento que permita responder la pregunta: Si el proceso está fuera de

control ¿cuál es la causa? (Maravelakis et al., 2002 e ISO 7870- 2, 2013).

En este trabajo, se utilizan datos de pesos de café en latas envasadas de la empresa “Central de Cooperativas Agrarias Cafetal Eras Cocla Ltda.” ubicada en la Provincia Constitucional del Callao.

## Materiales y Métodos

La investigación se aplicó en la empresa “Central de Cooperativas Agrarias Cafetal Eras Cocla Ltda.” ubicada en la Provincia Constitucional del Callao. La información utilizada se recolectó en el proceso de envasado. Se tomaron 26 muestras, cada una con tres observaciones correspondientes a los pesos del producto café en *sachets* (200 g).

### Equipos y materiales

Los datos fueron recolectados empleando equipos calibrados: balanza de precisión marca Ohaus Serie Explorer Pro, cronómetro marca RadioShack y un formato para recolección de datos. Softwares: “Minitab” versión 18 y “R” versión 3.5.2.

### Recolección de datos

La recolección de datos en la máquina envasadora Form Fill Seal Machine Modelo GL-250 para el producto café en sachets “Cafetal 200 g” se realizó a través de la toma de 26 muestras de tamaño 3 cada 10 minutos en el proceso de envasado en planta. La especificación establecida por el productor fue de  $200 \pm 2$  g.

### Diagrama de control de Shewhart

El diagrama de control para medias de procesos tiene como objetivos

observar y analizar mediante el uso de datos estadísticos la variabilidad del proceso de interés a través del tiempo, (Gutiérrez y De la Vara, 2004) y mostrar las fluctuaciones de las medias muestrales que se presentan dentro de estos límites. Si las medias muestrales caen dentro de los límites establecidos para un proceso (rango de aceptación), se dice que la variación que presenta el proceso solo es aleatoria. Pero si las medias muestrales exceden el límite superior de control (LSC) o bien, caen por debajo del límite inferior de control (LIC), entonces el proceso de producción o un servicio está fuera de control, y deberá corregirse.

Montgomery D. (2014), presenta las ecuaciones de las gráficas de control de Shewart para variabilidad (o amplitud) y la media:

$$LCL = D_3\bar{R} ; LC = \bar{R} ; UCL = D_4\bar{R}$$

Límite de control para la media:

$$LCL = \bar{x} - A_2\bar{R} ; LC = \bar{x} ; UCL = \bar{x} + A_2\bar{R}$$

Wu, Z. y Wang, QN. (1996), presentan la metodología Bootstrap para determinar gráficos de control cuando no se está seguro de la normalidad y muestras de tamaño menor a cuatro, como se detalla a continuación:

### Método Bootstrap

Para realizar la inferencia probabilística sobre cierto parámetro  $q$  de interés. Para ello, se toma una muestra  $X = (x_1, x_2, x_3, \dots, x_n)$  de tamaño  $n$  de una población empírica con distribución  $F$ .

Esta metodología considera los siguientes pasos:

- Construir una distribución de probabilidad empírica,  $\hat{F}(X)$  a partir de la muestra asignando una

probabilidad de  $1/n$  a cada punto,  $x_1, x_2, x_3, \dots, x_n$ . Esta es la función de distribución empírica (FDE) de  $X$ , la cual es el estimador no-paramétrico de máxima verosimilitud de la función de distribución de la población,  $F(X)$ .

- A partir de la FDE,  $\hat{F}(X)$ , se seleccionan  $B$  muestras aleatorias simples "con reemplazo" de tamaño  $n$
- Con los datos de cada una de las  $B$  muestras se calculan promedios:  $\bar{x}_i^*$  y los rangos  $R_i$ , donde  $i = 1, 2, \dots, B$ . La distribución de esos  $B$  valores se llama distribución empírica de Bootstrap.
- Usando la lista de  $B$ -promedios se calculan los límites de control con los métodos de aproximación a la normal y percentil respectivamente.

### Construcción del diagrama de control del rango R

Los límites Bootstrap del gráfico del rango  $R$  se definen según las ecuaciones:

$$LCL = \bar{R}^* - 3\hat{\sigma}_R^*$$

$$UCL = \bar{R}^* + 3\hat{\sigma}_R^*$$

El promedio de las amplitudes de las  $T$  muestras Bootstrap se obtiene siguiendo:

$$\bar{R}^* = \frac{R_1^* + R_2^* + \dots + R_T^*}{T}$$

Donde  $\hat{\sigma}_R^*$  es el error estándar Bootstrap calculado de la distribución muestral del

$$\text{Rango } R_n^*, \hat{F}(R^*)$$

### Construcción del diagrama de control del promedio $\bar{X}$

El promedio de las  $T$  medias Bootstrap se obtiene siguiendo:

$$\bar{X}^* = \frac{\bar{X}_1^* + \bar{X}_2^* + \dots + \bar{X}_T^*}{T}$$

Los límites Bootstrap del diagrama del promedio  $\bar{X}$  se definen según las ecuaciones:

$$LCL = \bar{\bar{X}}^* - 3\hat{\sigma}_{\bar{X}}^*$$

$$UCL = \bar{\bar{X}}^* + 3\hat{\sigma}_{\bar{X}}^*$$

Donde:  $\hat{\sigma}_{\bar{X}}^*$  es el error estándar Bootstrap calculado de la distribución muestral de la Media.

$$\bar{X}_n^*, \hat{F}(X^*)$$

**Construcción de los límites de control Bootstrap por el método percentil**

A continuación, se describe la construcción de diagramas de control de variables para muestras de tamaño “n” menores o iguales a tres, propuesto por Wu y Wang (1996).

**Construcción del diagrama de control del rango R:**

Sea  $R_1^*, R_2^*, \dots, R_T^*$ , la distribución Bootstrap de la distribución muestral del Rango  $R_n^*, \hat{F}(R^*)$ .

El intervalo percentil  $1-\alpha$  se define con  $\alpha/2$  y  $1-\alpha/2$  percentiles de  $\hat{F}(R^*)$ .

$$[R_b, R_p] \approx [R_T^{*(\alpha)}, R_T^{*(1-\alpha)}]$$

Donde:

$R_T^{*(\alpha/2)}$  es el  $100\alpha/2$  percentil de la distribución muestral de  $R^*$

$R_T^{*(1-\alpha/2)}$  es el  $100(1-\alpha/2)$  percentil de la distribución muestral de  $R^*$

Los límites Bootstrap del Gráfico del Rango R, se definen según las ecuaciones:

$$LCL = R_T^{*(\alpha/2)}$$

$$UCL = R_T^{*(1-\alpha/2)}$$

**Construcción del gráfico de control del promedio  $\bar{X}$ :**

Sea  $\bar{X}_1^*, \bar{X}_2^*, \dots, \bar{X}_T^*$  la distribución Bootstrap de la distribución muestral de Media  $\bar{X}_n^*, \hat{F}(\bar{X}^*)$ .

El intervalo percentil  $1-\alpha$  se define con  $\alpha/2$  y  $1-\alpha/2$  percentiles de  $\hat{F}(\bar{X}^*)$ .

$$[\bar{X}_b, \bar{X}_p] \approx [\bar{X}_T^{*(\alpha/2)}, \bar{X}_T^{*(1-\alpha/2)}]$$

Donde:

$\bar{X}_T^{*(\alpha/2)}$  es el  $100\alpha/2$  percentil de la función de densidad  $\bar{X}^*$ .

$\bar{X}_T^{*(1-\alpha/2)}$  es el  $100(1-\alpha/2)$  percentil de la función de densidad  $\bar{X}^*$ .

Los límites Bootstrap del gráfico del promedio  $\bar{X}$ , se definen según las ecuaciones:

$$LCL = \bar{X}_T^{*(\alpha/2)}$$

$$UCL = \bar{X}_T^{*(1-\alpha/2)}$$

**Medidas de eficacia de un diagrama de control**

**Longitud media de corrida (ARL)**

El ARL se define como el número promedio de puntos antes de que la carta de control de una señal de fuera de control sin que haya ocurrido algún cambio en el proceso. Es decir, esta señal de fuera de control se debe solo al azar del proceso. (Montgomery D., 2004).

$$ARL = \frac{1}{1-\beta}$$

Donde:

$1-\beta$  es igual a la *potencia* de la carta de control, la cual significa la **probabilidad** de rechazar correctamente los puntos que presentan una **condición fuera de control**

## Resultados y Discusión

### Construcción de los diagramas de control Shewhart $\bar{X}$ y R.

En esta etapa, se construye los diagramas de control  $\bar{X}$  y R para el peso del producto café Cafetal 200 gramos. Dos puntos que caen fuera de los límites de control  $\bar{X}_5 = 214.2$  y  $\bar{X}_3 = 195.7$  esto podría deberse a la existencia de causas atribuibles y que el proceso no es un sistema de causas constantes de variación. Se procedió a eliminar las muestras  $\bar{X}_5$

y  $\bar{X}_3$  para establecer diagramas que se encuentren bajo control. En la Tabla 1 se muestran los valores de los límites de los gráficos de control de Shewhart para el promedio ( $\bar{X}$ ) y rango (R) y en la Figura 1, se observa que todas las muestras caen dentro de los límites de control con una variación sostenida, además no hay señales de patrones no aleatorios, todo indica que solo existen causas inherentes al proceso; es decir, causas constantes productos del azar; por tanto, el proceso se encuentra en “estado de control”.

Tabla 1

Gráficos de control Shewhart  $\bar{X}$  y R para las medidas del peso de café.

N° de muestras	Promedio		Rango	
	LCL	UCL	LCL	UCL
26	0	14.17	195.9	207.1

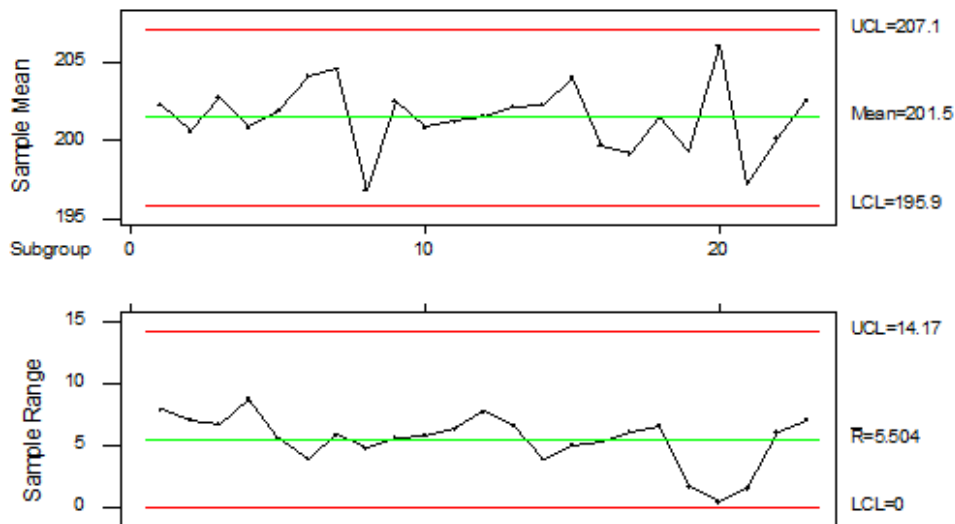


Figura 1. Gráficos de control Shewhart  $\bar{X}$  y R para las medidas del peso de café

### Construcción de los límites Bootstrap para los diagramas $\bar{X}$ y R para 1000 y 10000 muestras Bootstrap del proceso de envasado de sachets de café de 200 g

En esta etapa, se tomó en cuenta que la muestra de donde se extrajeron las muestras Bootstrap no presentó puntos

discordantes u outsiders ya que estos pudieran ser perpetuados, ocasionando una distorsión en la estimación Bootstrap de la distribución muestral.

A partir de datos de remuestreo se obtuvo la estadística descriptiva para el rango y para el promedio.

**Test de Normalidad, Estadística Descriptiva e intervalos para el rango en el Método Bootstrap**

Se realizó el test de normalidad y el análisis descriptivo para las 1000 y 10000 muestras Bootstrap de tamaño 3. La Figura 2 representa gráficamente el test de normalidad de los rangos de los pesos. Se observa que la distribución de los datos presenta un comportamiento aproximadamente normal, esto se verifica con la prueba de normalidad de Shapiro Wilk al obtener un p-value de 0.193, superior al nivel de significancia del 0.05, con lo cual se concluye que no existe suficiente evidencia estadística para rechazar la hipótesis de que los rangos de los pesos de las 1000 muestras Bootstrap de 200 g se ajustan a una distribución normal. En la Tabla 2 se presenta los

resultados de la estadística descriptiva aplicada a las muestras Bootstrap de sachets de café de 200 g.

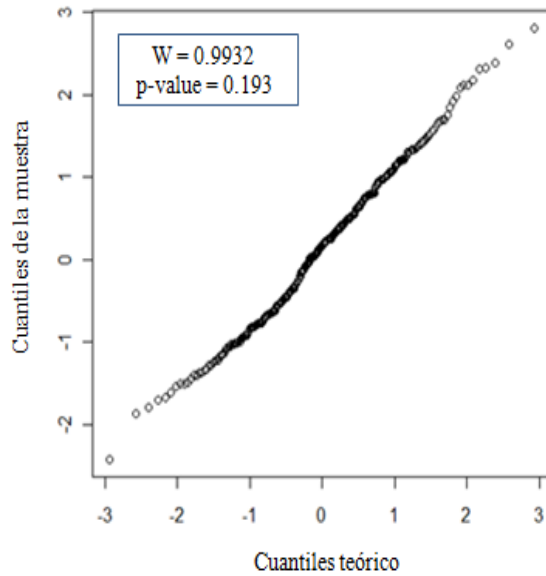


Figura 2. Normalidad para el rango del peso del envasado de 1000 y 10000 sachets Bootstrap de café de 200 g

Tabla 2  
*Estadística Descriptiva de los rangos de los pesos Bootstrap de sachets de café*

N° de muestras	Mín (g)	Q1 (g)	Promedio (g)	Q3 (g)	Max (g)	S (g)	CV (%)
1000	0.1	3.5	5.5745	7.8	12.1	2.6	47
10000	0	3.8	5.7671	7.8	5.767	2.58	45

Según el grado de dispersión, los datos presentados no son muy homogéneos de acuerdo con el coeficiente de variabilidad obtenido de 46.8 %. Los resultados muestran que hay cierta asimetría en los datos, las medidas de los rangos de las muestras Bootstrap varían entre 0.1 y 12.1 g. para simulaciones de 1000 muestras y es menor para simulaciones de 10 000

muestras entre 0 y 5.8 g.

**Obtención de los límites para los diagramas de control para le media ( $\bar{X}$ )**

En la Tabla 3 se presenta los resultados de la estadística descriptiva aplicada a las 1000 y 10000 muestras Bootstrap de sachets de café de 200 g.

Tabla 3  
*Estadística Descriptiva de los promedios de los de los pesos Bootstrap de sachets café*

N° muestras	Mín (g)	Q1 (g)	Promedio (g)	Q3 (g)	Max (g)	S (g)	CV (%)
1000	196.2	200.07	201.48	202.83	206.8	1.98	0.98
10000	195.43	200.13	201.48	202.83	206.6	1.9	0.94

En la Tabla 3, se observa un pequeño cambio en el promedio de los pesos, los datos presentados son muy homogéneos de acuerdo con el coeficiente de variabilidad obtenido de 0.98 % y 0.94 % respectivamente. Los resultados muestran que hay cierta asimetría en los datos, las medidas de los rangos de las muestras Bootstrap varían entre 196.20

y 206.8 g. para simulaciones de 1000 muestras y es mayor para simulaciones de 10 000 muestras entre 195.43 y 206.6 g.

En la Tabla 4, presenta los límites de control  $\bar{X}$  y R de la simulación de 1000 y 10000 muestras Bootstrap de tamaño  $n=3$ , con media 0 y desviación estándar 1 de una distribución normal.

Tabla 4

Límites de control del rango y promedio Bootstrap y Shewhart

	Shewhart	Bootstrap T =1000		Bootstrap T =10000	
		Normal	Percentil	Normal	Percentil
$LCL_R$	0	0	0.1	0	0.1
$UCL_R$	14.16	13.39	12.1	13.49	11.8
$LCL_{\bar{x}}$	195.87	195.54	196.57	195.8	196.57
$UCL_{\bar{x}}$	207.13	207.42	206.73	207.2	206.27

Según el grado de dispersión, los datos presentados son muy homogéneos de acuerdo con el coeficiente de variabilidad obtenido de 1.0 %. La comparación de los intervalos gráficos de control Normal .Bootstrap, el percentil Bootstrap y Shewhart se representa en las Figuras 5a y 5b con la simulación de 1000 y 10000

muestras. Se observa que la forma de la distribución de las medias de los pesos de las 1000 muestras Bootstrap de sachets de café de 200 g es moderadamente asimétrica hacia la derecha de acuerdo con lo señalado por la ISO (1976). Esto implica que la máquina envasadora se encuentra sobredosificando el café en los sachets de 200 g.

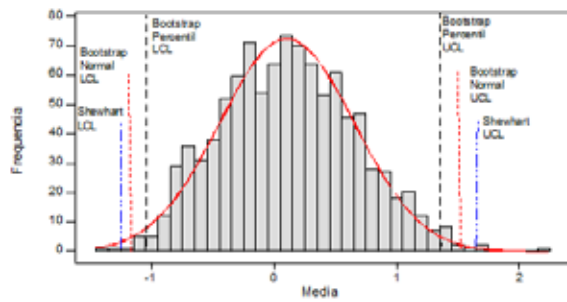


Figura 5a. Histograma de la media del peso de 1000 muestras Bootstrap de sachets de café

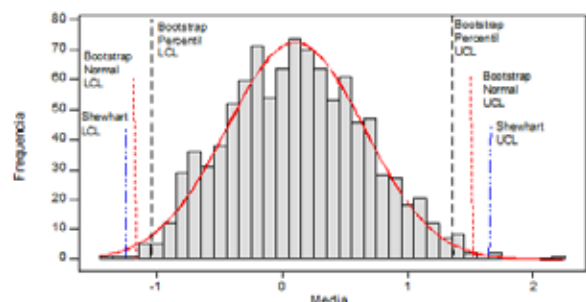


Figura 5b. Histograma de la media del peso de 10000 muestras Bootstrap de sachets de café

### Medidas de eficacia de un diagrama de control

Para medir la eficiencia de los gráficos de

control se calcula la longitud promedio de corrida (ARL).

Tabla 5

ARL para la media del Proceso  $H_0: \mu = 201.5$ ;  $H_1: \mu \neq 201.5$

Valores de la media	Shewhart	Bootstrap T =1000		Bootstrap T =10000	
		Normal	Percentil	Normal	Percentil
201.5	370.34	370.19	95.50	370.38	92.88
201.8	328.52	327.68	95.50	329.43	81.29
202.1	242.24	245.84	81.67	244.37	62.72
202.4	163.65	170.00	62.77	166.11	45.49
202.7	108.02	114.73	45.86	110.22	32.4
203.0	71.70	77.64	33.03	73.48	23.19
203.3	48.40	53.28	23.91	49.77	16.84
203.6	33.34	37.21	17.55	34.38	12.45
203.0	71.70	77.64	33.03	73.48	23.19
203.3	48.40	53.28	23.91	49.77	16.84
203.6	33.34	37.21	17.55	34.38	12.45

### Conclusiones

La metodología Bootstrap en la construcción de límites de control promedio-rango proporcionó límites de control más ajustados que los límites Shewhart, cuando se aumenta el número de simulaciones de 1000 a 10000 muestras T, incrementando la eficiencia. Los límites Bootstrap percentil tienden a ser más estrechos con respecto a los límites de control de promedio - rango

de Shewhart. Se puede afirmar, que la aplicación de la metodología Bootstrap permite obtener límites más ajustados e incrementar el poder de detección de los diagramas  $\bar{X}$  y R, así como, detectar cambios que se producen en el proceso. Por esta razón, se recomienda emplear los límites Bootstrap percentil cuando la población de donde procede la muestra tiene una distribución desconocida o cuando su distribución no se ajusta a la forma normal.

### Referencias

- Anderson, Sweeney D. y Williams T. (2004). Estadística para administración y economía, México, Thomson, 2004, pp. 825-830. También puede consultarse en R. Mason y D. Lind, Estadística para administración y economía, México, Alfa omega, 1998, pp. 911-915.
- Efron, B. y Tibshirani, R. J. (1993). An Introduction to the Bootstrap. First Edition. Chapman & Hall. New York. 436p.
- Gutiérrez, P. H. y De la Vara, S. R., (2004). Control estadístico de calidad y seis sigmas. México: McGraw-Hill Interamericana.
- ISO (International Organization for Standardization) ISO 2854 (1976). Statistical Interpretation of Data -- Techniques of Estimation and Tests Relating to Means and Variances. ISO. 46p.
- ISO. 1997. ISO 5479(1997). Statistical Interpretation of Data -- Tests



- for Departure from the Normal Distribution. ISO. 33p.
- ISO 7870-2(2013). Control charts -- Part 2: Shewhart control charts. ISO. 44p.
- Ledesma, R. (2008). Introducción al Bootstrap. *Tutorials in Quantitative Methods for Psychology*, 4(2), 51–60.
- Maravelakis P.E. and Panaretos J. (2002). Effect of estimation of the process parameters on the control limits of the univariate control chart for process dispersion. *Communications in Statistics*, 31 (3): 443–462. 2002.
- Montgomery D. (2014). *Control estadístico de la calidad*, Limusa-Wiley, México, 2004, 3a ed.
- Wu, Z. & Wang, QN. (1996). **BOOTSTRAP CONTROL CHARTS**, *Quality Engineering*, 9(1), 143-150, <https://doi.org/10.1080/08982119608919026>.

