

03-005

### **ON ESTIMATING THE CAPACITY INDEX OF A PROCESS BY USING SAMPLING STRATIFIED**

Moya Fernández, Pablo José  
Universidad de Granada

Currently, the statistical quality control is essential in the business world due to the competitive market in which it operates. The companies are required to have a suitable manufacturing process, hence, the importance of being accurate when evaluating the quality of a process. In practice, the capacity potential index and the located capability index are the common measures used for the analysis of the process capability. Traditionally, samples of products are selected under simple random sampling without replacement, and this is the information used for the determination of the various capability indices. This paper proposes a new process capability index, which is based upon the idea of stratified sampling. The proposed methodology assumes the concept of rational subgroups, which we also describe in this paper. Note that the concept of rational subgroups is similar to the principles of stratified sampling, and this implies that a new estimator with desirable properties can be defined. In addition, the suggested estimator can provide a more accurate view of reality.

**Keywords:** *Statistical Quality control; process capability analysis; rational subgroups; estimation method.*

### **ESTIMACIÓN DEL ÍNDICE DE CAPACIDAD DE UN PROCESO MEDIANTE MUESTREO ESTRATIFICADO**

En la actualidad, el control estadístico de calidad resulta fundamental en el ámbito empresarial debido al mercado competitivo en el que se opera. Tener un proceso de fabricación adecuado es básico para la empresa y, de ahí, la importancia de ser precisos en la evaluación de calidad del proceso. Hasta ahora se han utilizado para el análisis de capacidad de procesos algunos índices como el índice de capacidad potencial y el índice de capacidad localizado. Tradicionalmente se ha utilizado el muestreo aleatorio simple sin reemplazamiento para obtener las muestras de artículos en los que se basaban la determinación de estos índices de capacidad. En este trabajo se propone un nuevo índice de capacidad, basado en la técnica de muestreo estratificado con el que determinar la adecuación del proceso. Para esto será determinante el concepto de subgrupos racionales que se describe en este trabajo. En otras palabras, se expone que el concepto de subgrupo racional se adapta a los principios del muestreo estratificado y, por tanto, se consigue un nuevo estimador con mejores propiedades y que puede proporcionar una visión más fiel de la realidad.

**Palabras clave:** *Control estadístico de calidad; análisis de capacidad de un proceso; subgrupo racional; método de estimación*

Correspondencia: Pablo José Moya Fernández

pjmoyafernandez@gmail.com Agradecimientos: Juan Francisco

Muñoz; Encarnación Alvarez; Elisabeth Lahoz

## 1. Introducción

La calidad de los productos y/o servicios que se utilizan deberían ser uno de los aspectos más importantes a tratar por las empresas que los fabrican o bien los ofrecen. El control estadístico de calidad (Besterfield 1994; Chakraborti et al 2008; Montgomery 2009; Gutiérrez, 2013) es una de las numerosas herramientas que pueden utilizarse para evaluación y seguimiento de calidad de productos y servicios. En particular, los gráficos de control son unas técnicas incluidas dentro del control estadístico de calidad que permiten determinar si un determinado proceso de producción está fabricando o no productos dentro de unas determinadas especificaciones fijadas de antemano. En otras palabras, se trata de determinar si el proceso de producción se encuentra bajo control, en cuyo caso los productos cumplirán con las especificaciones establecidas. En caso contrario, se dice que el proceso está fuera de control.

Dentro del control estadístico de calidad se han definido una serie de índices que permiten conocer si el proceso puede considerarse apto o no para producir productos dentro de las especificaciones correspondientes. Algunos de estos índices son el índice de capacidad potencial y el índice de capacidad localizado (véase Montgomery 2009, pg. 351; Besterfield 1994, pg. 148).

Por otra parte, es importante tener en cuenta que tanto los gráficos de control como los índice de capacidad anteriormente comentados están basados en muestras, es decir, el control de calidad de productos se basa en las observaciones tomadas en los productos en una serie de muestras, las cuales se utilizan para la estimación de los parámetros asociados al proceso, estimación de los índices de capacidad, etc. El diseño muestral comúnmente utilizado para la selección de tales muestras es muestreo aleatorio simple sin reemplazamiento.

El esquema de muestreo utilizado para la selección de muestras puede producir resultados satisfactorios, es decir, las estimaciones podrían ser más precisas y tener sesgos más pequeños. Este hecho puede resultar fundamental para obtener gráficos de control e índices de capacidad más precisos que reflejen de la mejor forma lo posible el estado real del proceso de producción. Por ejemplo, es conocido que bajo determinadas condiciones el muestreo estratificado (Kish, 1965; Cochran 1977; Fernández y Mayor 1994; Särndal et al. 1992; Lohr 1999; Singh 2003) es más preciso que el tradicional muestreo aleatorio simple.

Los principales objetivos que se persiguen en este trabajo son: (i) definir el índice de capacidad localizado utilizando muestreo estratificado; (ii) evaluar empíricamente el comportamiento del índice de capacidad propuesto; y (iii) comparar tales resultados con el índice de capacidad localizado existente y que generalmente se utiliza en las empresas y procesos de fabricación. Para lograr estos objetivos, este trabajo presenta la estructura que se describe a continuación. En la Sección 2 tras una breve introducción al control de calidad y se presentan los conceptos clave de este trabajo: subgrupos racionales, índices de capacidad y muestreo estratificado. En la Sección 3 se plantea, como novedad, la estimación del índice de capacidad localizado mediante muestreo estratificado. En la Sección 4 se evalúa numéricamente el comportamiento del índice de capacidad propuesto mediante un estudio de Simulación Monte Carlo. Se comparan los resultados obtenidos con el comportamiento empírico del índice de capacidad utilizando muestreo aleatorio simple y se comprueba que pueden obtenerse resultados satisfactorios en términos de sesgo y eficiencia. Las principales conclusiones se describen en la Sección 5.

## 2. Índices de Capacidad y Muestreo Estratificado.

### 2.1. Introducción al control de calidad

El control de calidad puede definirse (véase, por ejemplo, Amor, 2005; Montgomery, 2009) como el conjunto de técnicas cuantitativas mediante las cuales se miden las características de calidad de un producto, se comparan con unas determinadas especificaciones fijadas de antemano, y se toman acciones correctivas apropiadas cuando hay muchos elementos que no cumplen los requisitos o condiciones fijadas.

En este trabajo se asume que la característica de calidad es una variable que se denota por  $X$ . Además se asume que dicha variable  $X$  tiene distribución Normal con media  $\mu$  y desviación típica  $\sigma$ , es decir,  $X \sim \mathcal{N}(\mu, \sigma)$ , donde  $\mu$  denota la media del proceso o característica y  $\sigma$  es la desviación típica. Destacar que generalmente los parámetros  $\mu$  y  $\sigma$  son desconocidos por lo que se utiliza la estimación puntual de los mismos. En este caso se seleccionan  $m$  muestras de tamaño  $n$  mediante muestreo aleatorio simple sin reemplazamiento cuando se tiene certeza de que el proceso se encuentra bajo control. En la Tabla 1 se recoge la notación utilizada para denotar las observaciones recogidas en las  $m$  muestras, donde  $x_{ij}$  denota el valor de  $X$  para la  $i$ -ésima muestra y la  $j$ -ésima observación.

**Tabla 1. Observaciones obtenidas en  $m$  muestras de tamaño  $n$  de la característica  $X$**

| Muestra | Observaciones |          |   |          |
|---------|---------------|----------|---|----------|
| 1       | $x_{11}$      | .        | . | $x_{1n}$ |
| .       | .             | $x_{ij}$ | . | .        |
| .       | .             | .        | . | .        |
| $m$     | $x_{m1}$      | .        | . | $x_{mn}$ |

Dado el conjunto de observaciones especificado en la Tabla 1, la estimación de la media poblacional ( $\mu$ ) viene dada por la expresión

$$\hat{\mu} = \bar{\bar{x}} = \frac{1}{m} \sum_{i=1}^m \bar{x}_i \quad (1)$$

donde  $\bar{x}_i$  es la media muestral de muestra  $i$ , definida como  $\bar{x}_i = \frac{1}{n} \sum_{j=1}^n x_{ij}$ .

Por su parte, la desviación típica del proceso ( $\sigma$ ) puede ser estimada mediante diferentes métodos basados en: (i) las cuasi desviaciones típicas, (ii) los rangos de las muestras, y (iii) las cuasi desviación típica combinada de la muestra. Los estimadores se determinan como sigue:

$$\hat{\sigma}_S = \frac{\bar{S}}{c_4[n]}; \hat{\sigma}_R = \frac{\bar{R}}{d_2[n]}; \hat{\sigma}_p = \frac{S_p}{c_4[1 + m(n-1)]} \quad (2)$$

donde  $\bar{S} = m^{-1} \sum_{i=1}^m S_i$  es la media de las cuasi desviación típicas de las muestras, definidas como  $S_i = \left( \frac{1}{n-1} \sum_{j=1}^n (x_{ij} - \bar{x}_i)^2 \right)^{1/2}$ ;  $\bar{R} = m^{-1} \sum_{i=1}^m R_i$  es la media de los rangos  $R_i = \max(x_{ij}) - \min(x_{ij})$  y  $S_p = \left( \frac{1}{m} \sum_{i=1}^m S_i^2 \right)^{1/2}$ ;  $c_4[\cdot]$  y  $d_2[\cdot]$  son constantes basadas en los tamaños muestrales y las cuales están definidas, por ejemplo, por Chen (1997). Además, el

Apéndice VI de Montgomery (2009) contiene tablas para estas constantes y para varios valores de  $n$ .

El control de calidad distingue varios tipos de límites:

- Límites de tolerancia natural: son consecuencia de la variabilidad natural del proceso. Vienen dados por los extremos del intervalo en el que toma valores la variable del proceso,  $X$ . Asumiendo el método USA, los límites de tolerancia natural de un proceso suelen venir dados por  $\mu \pm 3\sigma$ .
- Límites de control: Son los límites de tolerancia natural basados en una muestra de tamaño  $n$ . 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  $\mu \pm 3\sigma/\sqrt{n}$ .
- Límites de Especificación: son límites dados sin tener en cuenta la variabilidad del proceso. Estos límites son impuestos por agentes externos. Se distingue entre Límite Inferior de Especificación (LIE) y Límite Superior de Especificación (LSE).

El siguiente concepto relevante en el contexto del control estadístico de calidad que se utiliza en este trabajo es el definido por Shewhart (1931) como subgrupos racionales. Este concepto está relacionado con una recopilación de datos de la muestra óptima, en el sentido de obtener resultados y conclusiones más fiables. Según Montgomery (2009), las muestras deberían seleccionarse de modo que si en el proceso existen causas asignables, la posibilidad de que existan diferencias entre los subgrupos o muestras sea máxima, mientras que la probabilidad de diferencias debido a estas causas asignables dentro de un subgrupo sea la mínima posible. Este concepto de subgrupos racionales está estrechamente relacionado con el esquema o escenario apropiado para el uso de un muestreo estratificado. Esta relación directa entre subgrupos racionales en el control estadístico de calidad y el muestreo estratificado ha sido el punto de partida o la motivación para este trabajo.

Por tanto, queda de manifiesto la importancia del concepto de subgrupos racionales, puesto que permite una selección de muestra que incrementa la información del control estadístico de la calidad. Así mismo, es necesario un cuidadoso estudio del proceso de producción para la elección de la estrategia a seguir en la elaboración de los subgrupos racionales.

## 2.2. Análisis de Capacidad

El análisis de capacidad de un proceso tiene como objetivo evaluar la aptitud de un proceso para producir unidades dentro de los límites de especificación. Este análisis de capacidad de un proceso se puede efectuar mediante el índice de capacidad y el índice capacidad localizado.

El índice de capacidad potencial de un proceso se define como el cociente entre la amplitud de los límites de especificación y la amplitud de los límites de tolerancia natural del proceso. Matemáticamente, el índice de capacidad potencial,  $C_p$ , de proceso se obtiene mediante la siguiente expresión

$$C_p = \frac{LSE - LIE}{6\hat{\sigma}}, \quad (3)$$

donde  $6\hat{\sigma} = LST - LIT$ . Con el índice de capacidad potencial de un proceso se compara la variabilidad natural del proceso con el rango de valores permitidos en el proceso y definidos a través de los límites de especificación, dando lugar a una de las siguientes situaciones: (i)  $C_p = 1$ , el proceso está marginalmente apto para cumplir con los límites de especificación; (ii)  $C_p < 1, LSE - LIE < 6\hat{\sigma}$ ; el proceso no es capaz de cumplir con los límites de

especificación; y (iii)  $C_p > 1, LSE - LIE > 6\hat{\sigma}$ ; el proceso se considera apto para cumplir con los límites de especificación.

En resumen, cuanto mayor sea el índice de capacidad potencial de un proceso, mayor será la capacidad del proceso para cumplir con las especificaciones. En la práctica, las empresas consideran una cota mínima del índice de capacidad dada por  $C_p \geq 4/3$ , y establecen como meta la obtención de un nivel  $C_p \geq 5/3$ . Por último, destacar que índice de capacidad potencial es adimensional, y esto permite comparar entre distintos procesos de forma que se pueda conocer el proceso que tiene una mayor capacidad.

El índice de capacidad localizado se utiliza cuando la línea central del gráfico de control no coincide con el punto medio de los límites de especificación y su interpretación es similar al anterior índice. De esta manera, el índice de capacidad localizado se define como:

$$C_{pk} = (1 - k)C_p, \quad (4)$$

donde  $k$  es la distancia relativa entre la media del proceso ( $\hat{\mu} = \bar{x}$ ) y el punto medio de los límites de especificación ( $\mu_0 = (LIE + LSE)/2$ ), es decir:

$$k = \frac{|\mu_0 - \hat{\mu}|}{(LSE - LIE)/2} \quad (5)$$

De nuevo se presentan tres posibilidades: (i)  $k=1$ , la distancia entre la media del proceso y el punto medio de los límites de especificación es nula; (ii)  $k>1$ , que la distancia entre la media del proceso y el punto medio de los límites de especificación es superior que la mitad de la amplitud del intervalo, lo que implica que el proceso producirá un número excesivamente alto de artículos defectuosos; y (iii)  $k<1$ , la distancia entre la media del proceso y el punto medio de los límites de especificación es inferior que la mitad de la amplitud del intervalo, lo que indica que la media del proceso se encuentra entre los límites de especificación pero esto no significa que el proceso esté produciendo un número óptimo de artículos conforme a las especificaciones.

### 2.3. Muestreo Estratificado

El muestreo estratificado es una técnica de muestreo muy conocida y utilizada en por los organismos oficiales de estadística así como otras agencias u organismos encargados de realizar estudios y que persigan el objetivo de obtener estimaciones más precisas. En el muestreo estratificado se divide la población en subpoblaciones o estratos, y en cada uno de estos subgrupos de individuos se selecciona una muestra, de forma que la muestra total de individuos estará formada por la unión de todas las muestras obtenidas en los distintos estratos. Cuando se aplica un muestreo aleatorio simple en la selección de muestras en cada estrato, al muestreo estratificado se le denomina muestreo aleatorio estratificado.

Para que el muestreo estratificado pueda obtener estimaciones más precisas que el tradicional muestreo aleatorio simple, los estratos deben ser heterogéneos entre sí, mientras que los individuos pertenecientes a un mismo estrato deberían tener valores homogéneos en la variable objeto de estudio. Se observa que estas condiciones de heterogeneidad entre estratos y homogeneidad dentro de estratos son las que se persiguen con la idea de subgrupos racionales introducidas en el campo del control estadístico de calidad. Se pretende utilizar un esquema de muestreo estratificado en el problema de estimación del índice de capacidad de un proceso, dado que dicho índice depende del parámetro desconocido  $\mu$ , y por tanto se requiere de una estimación precisa de dicho parámetro.

### 3. Estimador propuesto de Índice de Capacidad usando Muestreo Estratificado.

El objetivo principal de la sección anterior ha sido introducir, de forma muy breve, la notación y conceptos básicos en el campo del control estadístico de calidad así como el conocido muestreo estratificado. Una vez conocidos estos conceptos, a continuación se resumen las condiciones y escenarios que se asumen en este trabajo:

(C1) La característica de calidad se mide mediante una variable  $X$ .

(C2) La variable  $X$  tiene distribución Normal con parámetros  $\mu$  y  $\sigma$ , es decir,  $X \sim \mathcal{N}(\mu, \sigma)$ .

(C3) Se usan  $m$  lotes de productos para la estimación de la media poblacional  $\mu$ .

(C4) El tamaño de cada lote de productos se denotará como  $N_i$ , es decir, el  $i$ -ésimo lote contiene  $N_i$  productos.

(C5) En cada lote se selecciona una muestra de productos mediante muestreo aleatorio simple sin reemplazamiento.

(C6) Las muestras obtenidas en los distintos lotes de productos tienen el mismo tamaño muestral, el cual se denotará como  $n$ , siguiendo la notación habitual dentro del contexto de control estadístico de calidad.

En esta sección se propone estimar el índice de capacidad localizado utilizando la técnica de muestreo denominada muestreo estratificado. En líneas generales, este proceso se basa en seleccionar  $m$  muestras, de tamaño  $n$  cada una de ellas, correspondientes a varias poblaciones objetivo (lotes) de ese proceso, medir la característica de calidad objeto de estudio, la cual se supone distribuida de forma normal, y aplicar los índices que analizan la capacidad del proceso. La idea principal es agregar el muestreo estratificado en la estimación del parámetro desconocido  $\mu$ , de manera que el estimador resultante sea más eficiente y con sesgos más pequeños.

Cuando se presentó el concepto de subgrupos racionales se hizo hincapié en que estos subgrupos deberían, ante causas atribuibles en los procesos, hacer máximas las diferencias entre grupos y mínimas las diferencias dentro de ellos. En consecuencia, por este motivo es razonable pensar que puede resultar favorable utilizar el muestreo estratificado tomando cada uno de estos subgrupos racionales de manera similar a un estrato. Así, los aspectos que en un principio se presentaban como desventajas atribuibles a las peculiaridades técnicas en el muestreo estratificado no tienen repercusión en este desarrollo, puesto que tanto la forma de seleccionar los estratos como el número de los mismos vienen dado por los subgrupos racionales.

Además, este nuevo índice calculado con el nuevo estimador del parámetro  $\mu$ , otorga diferentes pesos a los estratos l subgrupos racionales en función del tamaño de éstos. En consecuencia, cuanto mayor sea el tamaño del estrato l subgrupo racional más influencia tendrá para la determinación del estimador.

En conclusión, el nuevo estimador de la media, utilizando muestreo estratificado, se define como:

$$\hat{\mu}_{st} = \bar{x}_{st} = \sum_{i=1}^m W_i \bar{x}_i = \sum_{i=1}^m \frac{N_i}{N} \cdot \bar{x}_i \quad (6)$$

donde  $N_i$  es el tamaño correspondiente al lote de la muestra  $i$ -ésima,  $N$  el tamaño total de los lotes ( $N = \sum_{i=1}^m N_i$ ) y  $W_i$  la ponderación que se otorga en función al tamaño relativo de la muestra.

En la Tabla 2 se recoge la notación utilizada para denotar las observaciones recogidas en las  $m$  muestras. En este caso, se añade una columna con el tamaño de los lotes de donde se obtiene la muestra.

**Tabla 2. Observaciones obtenidas en  $m$  muestras de tamaño  $n$  de la característica  $X$  correspondientes al estimador en muestreo estratificado**

| Muestra | Observaciones |          |          |   |          | $N_i$ |
|---------|---------------|----------|----------|---|----------|-------|
| 1       | $X_{11}$      | $X_{12}$ | .        | . | $X_{1n}$ | $N_1$ |
| 2       | $X_{21}$      | $X_{22}$ | .        | . | $X_{2n}$ | $N_2$ |
| 3       | .             | .        | .        | . | .        | .     |
| .       | .             | .        | $X_{ij}$ | . | .        | .     |
| .       | .             | .        | .        | . | .        | .     |
| .       | .             | .        | .        | . | .        | .     |
| $m$     | $X_{m1}$      | $X_{m2}$ | .        | . | $X_{mn}$ | $N_m$ |

Finalmente, el índice de capacidad localizado, utilizando muestreo estratificado, viene dado por la siguiente expresión:

$$C_{pk}^{st} = (1 - k_{st})C_p \quad (7)$$

donde  $k_{st}$  es la distancia relativa entre la media estimada del proceso a través del nuevo método ( $\hat{\mu}_{st} = \bar{x}_{st}$ ) y el punto medio de los límites de especificación ( $\mu_0 = (LIE + LSE)/2$ ), a saber:

$$k_{st} = \frac{|\mu_0 - \hat{\mu}_{st}|}{\frac{LSE - LIE}{2}} \quad (8)$$

El índice de capacidad localizado basado en muestreo estratificado se interpreta de igual manera que el tradicional índice de capacidad localizado, ya que sólo trata de tener una mayor precisión de la capacidad del proceso sin perder su significado.

#### 4. Estudio de Simulación

Asumiendo las condiciones y escenarios descritos en la Sección 3 del trabajo, en esta Sección 4 se ha llevado a cabo un estudio de simulación Monte Carlo con el objetivo de evaluar empíricamente el comportamiento de los estimadores propuestos  $\hat{\mu}_{st}$  y  $C_{pk}^{st}$ . Tales resultados se compararán con los resultados empíricos obtenidos a través de los estimadores tradicionales existentes en la literatura y también definidos en este trabajo.

En este estudio de simulación se ha tenido en cuenta el concepto de subgrupos racionales, de tal manera que las muestras se obtienen de lotes con diferentes medias. En particular, se asume que los productos en los distintos lotes se generan mediante una distribución Normal con una media dentro del intervalo  $(\mu - \epsilon, \mu + \epsilon)$ , donde se considera  $\mu = 10$ , y donde  $\epsilon$  es una determinada constante. Como desviación típica del proceso se considera  $\sigma = 1$ . En este trabajo, se consideran los valores:

$$\epsilon = (0, 0.05, 0.1, 0.2, 0.5, 1, 2) \quad (9)$$

para esta constante que determina el rango de valores para la media del proceso en los distintos lotes. Con el fin de conocer la diferencia, en términos relativo, entre los extremos del intervalo y la media del proceso, se define la variación relativa

$$V = \frac{(\mu + \epsilon) - \mu}{\mu} = \frac{\epsilon}{\mu} = \frac{\epsilon}{10}, \quad (10)$$

cuyos valores en esta simulación son: 0%; 0,5%; 1%; 2%; 5%; 10%; 20%. Por ejemplo,  $V = 10\%$  indica una variación del 10% entre los extremos del intervalo y el valor de  $\mu$ . El objetivo de esta dispersión en cuanto a los valores de la media del proceso es poder observar la evolución de los estimadores y los índices de capacidad a medida que las muestras seleccionadas de los distintos lotes sean más heterogéneas entre sí, lo cual debería favorecer el comportamiento del estimador propuesto en comparación con el estimador tradicional. Los tamaños de los lotes (los valores  $N_i$ ) se han tomado entre 50 y 145.

El sesgo relativo empírico (SR) y el error cuadrático medio relativo empírico (ECMR) son las medidas empíricas usadas para comparar los distintas estimaciones de  $\mu$  y del índice de capacidad. Para el caso de la media, por ejemplo, estas medidas se definen como:

$$SR = \frac{E[\tilde{\mu}] - \mu}{\mu} \quad ; \quad ECMR = \frac{\sqrt{ECM[\tilde{\mu}]}}{\mu}, \quad (11)$$

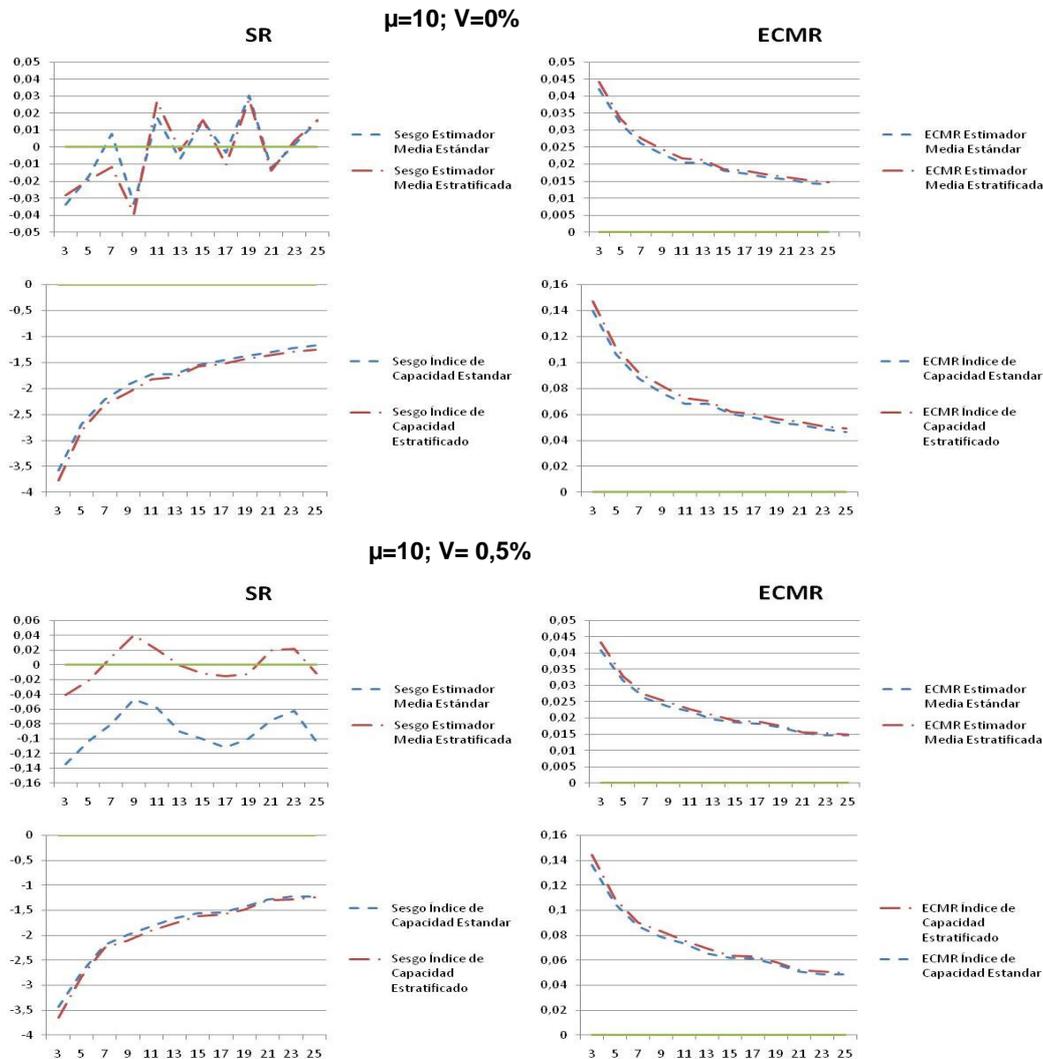
donde  $\tilde{\mu}$  es un determinado estimador de  $\mu$ , y  $E[\cdot]$  y  $ECM[\cdot]$  son, respectivamente, la esperanza empírica y el error cuadrático medio empírico basados en  $R = 1000$  simulaciones realizadas, es decir,

$$E[\tilde{\mu}] = \frac{1}{R} \sum_{i=1}^R \tilde{\mu}(i) \quad ; \quad ECMR[\tilde{\mu}] = \frac{1}{R} \sum_{i=1}^R (\tilde{\mu}(i) - \mu)^2, \quad (12)$$

donde  $\tilde{\mu}(i)$  denota el valor de la estimación  $\tilde{\mu}$  para la  $i$ -ésima simulación realizada. Para el caso de la estimación del índice de capacidad, estas medidas se definen de forma similar.

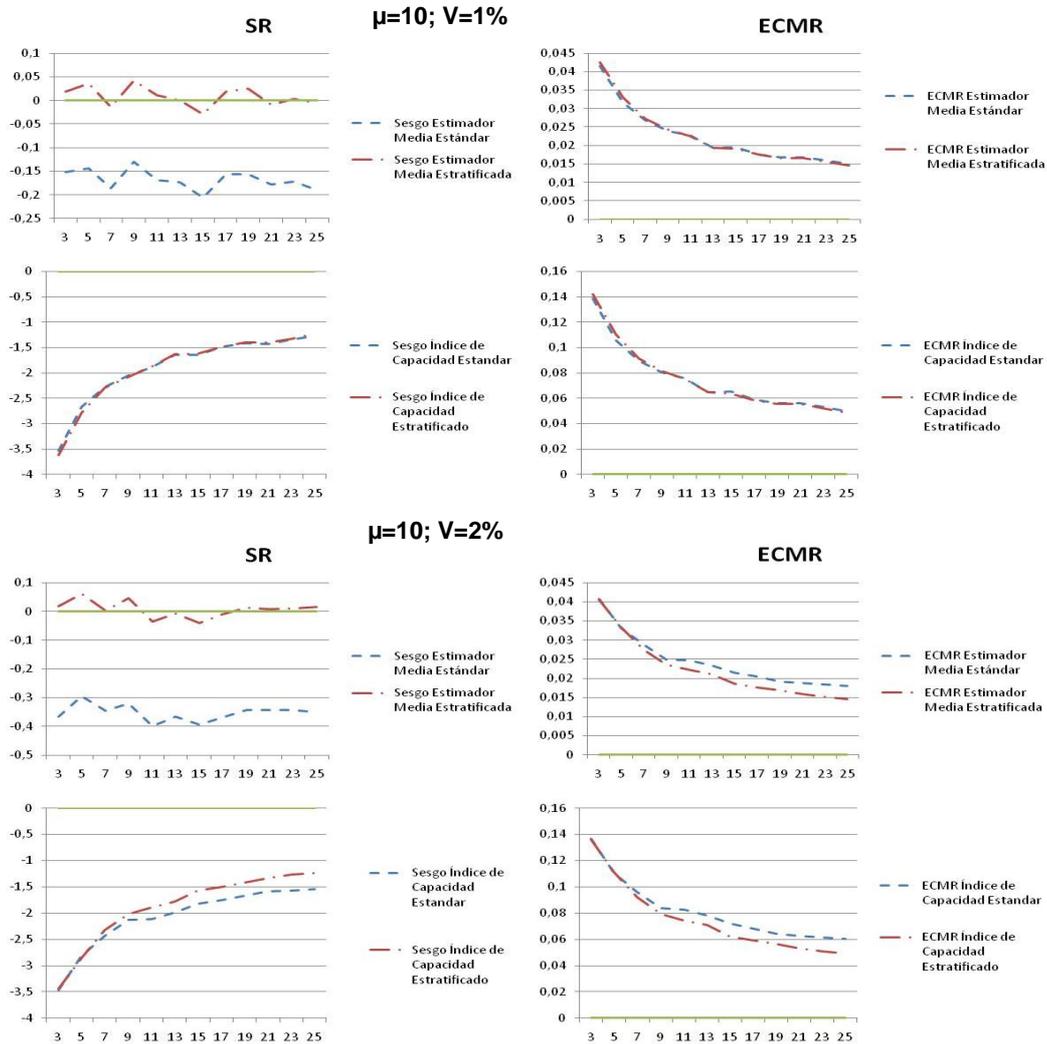
Destacar que las medidas SR y ECMR son muy comunes en los problemas de comparación de precisión de estimadores. Por ejemplo, éstas han sido usadas por Chen y Sitter (1999), Deville y Särndal (1992), Rao et al. (1990), Silva y Skinner (1995), etc. Los resultados derivados de este estudio de simulación se muestran en las Figuras 1, 2 y 3.

**Figura 1: Valores de SR y ECMR (en porcentaje) de estimadores de la media del proceso y del índice de capacidad localizado. Valores de V del 0% y del 0,05%.**



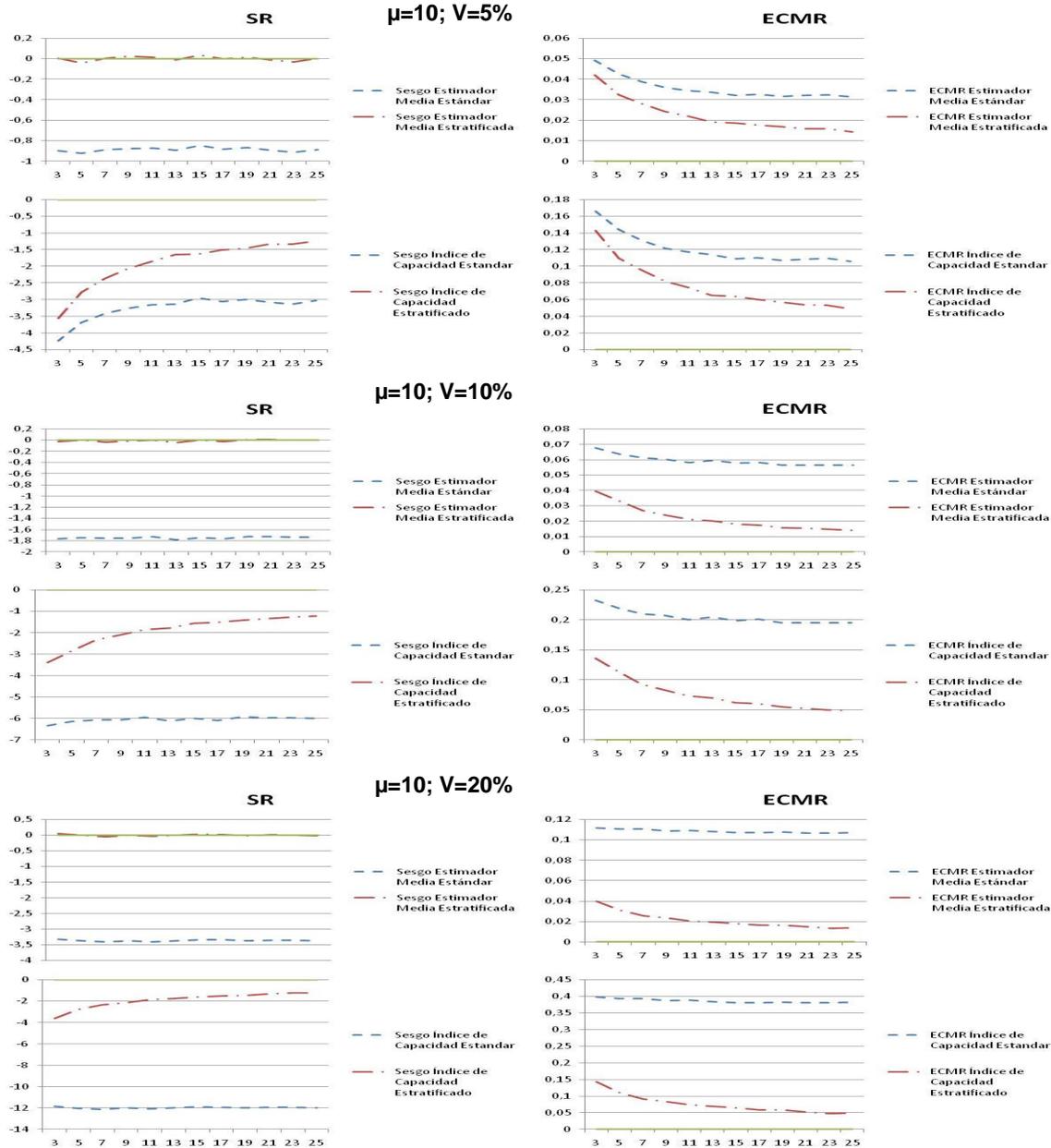
En la Figura 1 se observa que los sesgos de ambos estimadores (estándar y el propuesto usando muestreo estratificado) de la media del proceso son insignificantes, puesto que, en valor absoluto, nunca se supera el 1%. Cuando  $V = 0.5\%$ , se puede observar que el estimador estándar tiene, ligeramente, un sesgo negativo. Respecto al problema de la estimación del índice de capacidad localizado, se observa que se pueden obtener sesgos negativos de hasta el 4% y en muestras pequeñas. A medida que se aumenta el tamaño de la muestra, estos sesgos van disminuyendo hasta llegar aproximadamente al 1% en términos absolutos. En relación a las eficiencias de los estimadores en términos de ECMR, se observa que ambos estimadores tienen resultados similares en la estimación de ambos parámetros. Como es razonable, se obtienen valores más eficientes a medida que aumenta el tamaño de muestra.

**Figura 2: Valores de SR y ECMR (en porcentaje) de estimadores de la media del proceso y del índice de capacidad localizado. Valores de V del 1% y del 2%.**



En la Figura 2 se observa que los sesgos de ambos estimadores siguen estando por debajo del 1%, por lo que no representan un problema. Sin embargo, comienza a apreciarse una leve diferencia entre los sesgos de los estimadores. De esta manera, cuando  $V = 1\%$ , el estimador estándar presenta sesgos negativos de aproximadamente 0,15% mientras el estimador con muestreo estratificado sesgos no superiores al 0,05%. Cuando  $V = 2\%$ , estas diferencias entre los sesgos de estimadores siguen siendo leves pero algo más apreciable, el estimador con muestreo estratificado sigue obteniendo los mismos sesgos aproximadamente, mientras que los sesgos del estimador estándar se sitúan prácticamente en el 0,4%, evidentemente también un sesgo insignificante. Centrando la atención en los sesgos de los índices de capacidad usando ambos estimadores, se observa que los comportamientos de éstos son similares a los anteriores. Sin embargo se aprecia que a medida que aumenta  $V$  la mejora del índice de capacidad estándar es menor que la mejora del índice de capacidad con muestreo estratificado. Por su parte, los ECMR presentan un comportamiento similar aunque más sutil. Cuando  $V = 1\%$ , las diferencias entre ambos estimadores son prácticamente imperceptibles en las Figuras utilizadas en este trabajo. Sin embargo, cuando  $V = 2\%$ , comienza a entrecerse una leve diferencia entre ambos.

**Figura 3: Valores de SR y ECMR (en porcentaje) de estimadores de la media del proceso y del índice de capacidad localizado. Valores de V del 5%, 10% y del 20%.**



Por último, en la Figura 3 se pueden observar los sesgos y la eficiencia de los estimadores para  $V = 5\%$ ,  $10\%$  y  $20\%$ . El comportamiento de los sesgos del estimador usando muestreo estratificado es similar a los casos anteriores, es decir, sesgo próximo al 0%. Sin embargo, a medida que  $V$  aumenta, el comportamiento del estimador estándar es peor. Sus sesgos han aumentado situándose sobre el 0,9%, 1,8% y cerca del 3,5% respectivamente. Por su parte, los sesgos de los índices de capacidad muestran un comportamiento parecido. En el primer caso ( $V = 5\%$ ), ambos índices mejoran al aumentar el tamaño de muestra, pero el índice de capacidad estándar presenta, aproximadamente, sesgos negativos un punto porcentual mayor que el índice de capacidad con muestreo estratificado. Respecto a la eficiencia (valores de ECMR) los distintos estimadores reflejan comportamientos similares. Se puede observar cómo a medida que aumenta  $V$  y aumentan los tamaños de muestra, el estimador propuesto es más eficiente que el estimador estándar.

## 5. Conclusiones

En este trabajo se ha propuesto un nuevo índice para analizar la capacidad de un proceso. Este indicador consiste en utilizar muestreo estratificado como técnica de muestreo con el objetivo de hacer una estimación más precisa de la media del proceso,  $\mu$  y, así, obtener un índice que pueda tener una mayor precisión o sesgos más pequeños. Destacar que para que el estimador propuesto pueda tener un mejor comportamiento, es preciso que las muestras de productos se seleccionen siguiendo el concepto de subgrupos racionales definido dentro del contexto del control estadístico de calidad.

Se ha llevado a cabo un estudio de simulación Monte Carlo basado en 1000 replicas. Para contrastar si este índice de capacidad estratificado es más preciso que los índices tradicionales se ha tenido en cuenta dentro del estudio de simulación el concepto de subgrupos racionales. Para ello se ha considerado que cada lote presentaba una media que no era común para todos los lotes. Se asume que el valor de la media del proceso puede tomar cualquier valor dentro de un intervalo fijado. Los resultados obtenidos en este estudio de simulación muestran que conforme se incrementa la amplitud de este intervalo en el que se encuentra la media de los lotes, más preciso es el nuevo índice de capacidad propuesto. Dicho de otra forma, tanto los sesgos relativos (en valores absolutos) como los errores cuadráticos medios relativos muestran valores más pequeños en comparación con los estimadores habituales que se utilizan en la práctica.

En resumen, estos resultados revelan la clave que permite aplicar el muestreo estratificado en el análisis de procesos: el concepto de subgrupo racional. Cuando se considera un valor  $V = 0\%$ , la selección de las muestras de cada lote se realiza sin tener en consideración la existencia de subgrupos racionales. Bajo esta situación, la principal conclusión es que el estimador tradicional de la media del índice de capacidad presenta un comportamiento similar, e incluso en algunas ocasiones mejor, que los estimadores propuestos. Sin embargo, a medida que al aumentar el intervalo de variación donde se puede situar la media de cada lote, es decir, a medida que los subgrupos racionales son más heterogéneos entre sí, mejor comportamiento presentan los estimadores propuestos. Estos resultados se pueden consultar en la Figura 3 donde se representan los resultados para valores elevados de  $V$ . Subrayar que estos resultados son coherentes con la teoría existente del muestreo estratificado.

Por último y como conclusión final, recalcar que el índice de capacidad usando muestreo estratificado puede ser más preciso y tener sesgos más pequeños en comparación con los índices de capacidad tradicionales cuando las muestras de los lotes que se selecciona se hacen a través del concepto de subgrupos racionales establecidos en el contexto del control estadístico de calidad.

## BIBLIOGRAFÍA

- Amor R., Martínez A., Olmo M., (2005). *Control Estadístico de Calidad*. ed. Grupo Editorial Universitario.
- Besterfield, D. (1995). *Control de Calidad*. PRENTICE HALL HISPANOAMERICANA S.A.
- 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.
- 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.
- 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.
- Cochran W.G. (1977). *Sampling Techniques*. New York: John Wiley.
- Deville, J. C., Särndal, C. E. (1992) Calibration estimators in survey sampling. *J Am Statist Assoc*, 87; 376-382.
- Fernández García, F.R. y Mayor Gallego, J.A. (1994) *Muestreo en Poblaciones Finitas: Curso Básico*. P.P.U., Barcelona.
- Gutiérrez Pulido, Humberto (2013) *Control Estadístico de la Calidad y Seis Sigmas*. Ed. México: McGrawHill.
- Kish, L. (1965) *Survey Sampling*. New York. John Wiley.
- Lohr, S. L. (1999). *Sampling: Design and Analysis*. Pacific Grove, CA: Duxbury Press.
- Montgomery, D.C. (2009). *Statistical quality control. A modern introduction. 6th ed*. New York, Wiley.
- Nascimento-Silva, P. L. D., Skinner, C. J. (1995). Estimating distribution functions with auxiliary information using Post-stratification. *Journal of Official Statistics*, 11; 277-294.
- 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.
- Särndal, C. E., Swensson, B., Wretman, J. H. (1992). *Model Assisted Survey Sampling*. 3rd ed. New York: Springer-Verlag.
- Shewhart, W.A. (1931). *Economic control of quality of manufactured product*. Milwaukee, ASQC Quality Press.
- Singh, S. (2003). *Advanced Sampling Theory with Applications: How Michael Selected Amy*. Dordrecht, The Netherlands: Kluwer.

