



## Transformación De Datos Para El Control Estadístico De Procesos

C. A. Mendoza García.  
E. A. Ruelas Santoyo  
Departamento de Ingeniería Industrial  
Instituto Tecnológico Superior De Irapuato

### Resumen

El principal objetivo de este proyecto es encontrar una forma eficaz de analizar los datos de un proceso y que los resultados obtenidos sean confiables, sin importar la distribución que estos sigan. Se diseñó una herramienta estadística que es capaz de analizar datos y determinar su normalidad. En caso de que los datos sigan una distribución normal, automáticamente se determinan los índices de capacidad  $C_p$  y  $C_{pk}$ . En caso contrario, que los datos sean no normales se realiza la transformación de dichos datos para así poder llevar a cabo un análisis verídico. Se realizó un estudio de las diferentes técnicas de análisis de datos a fin de identificar la óptima y en base a esto se determinó que la prueba de Box- Cox es la transformación que resulta más completa y se determinó esta como la base de la herramienta estadística.

Palabras clave: control estadístico de procesos, distribución no normal, transformación estadística.

### Introducción

Para toda organización es sumamente importante generar utilidades mediante el aumento de ventas y disminución de costos. Para lo cual, es fundamental el cumplimiento de las especificaciones del producto a manufacturar y una distribución normal en el proceso de producción.



La distribución de un proceso depende de las medidas de tendencia central y dispersión que este registre, basado en la producción y el tamaño de muestra que se encuentre bajo estudio (Pulido, H. y de la Vara R., 2008). Esta muestra debe ser aleatoria y de acuerdo al Teorema de Límite Central constar de mínimo 30 elementos para ser válida (Berenson, Levine y Krehbiel, (2001)). Para comprobar la normalidad de un proceso existen diferentes pruebas de control estadístico, como la prueba Anderson-Darling. Esta prueba dicta dos hipótesis. La hipótesis nula indica normalidad en los datos y la hipótesis alternativa indica que los datos no son normales. Si el p-valor es menor al nivel de significancia establecido, se rechaza la hipótesis nula y se entiende que los datos no son normales, según (Salgado, 2019).

Sin embargo, existen ocasiones en las que estas pruebas no son aplicadas en el análisis de procesos y estos se analizan suponiendo normalidad en los datos, dando como consecuencia resultados inválidos para estudio y análisis. Los datos no normales, se refiere a medidas o datos fuera de los límites de control de calidad. Estos datos no pueden ser aplicados directamente a alguna herramienta de control de proceso, si no que primero deben pasar por un proceso de transformación a datos normales. Hoy en día no existe una herramienta que directamente identifique la normalidad de los datos recopilados y en caso de no normalidad los transforme adecuadamente, con el procedimiento necesario, ese es el objetivo principal de esta investigación, crear una herramienta para la detección de no normalidad y transformación de esta por medio del método de Box Cox que nos permite normalizar los datos a través de la variable lambda con el objetivo de obtener datos normales para llevar a cabo una correcta interpretación de los cálculos estadísticos. Con el propósito de brindar una metodología que detecte el tipo de distribución, llevando a cabo correctamente los cálculos de media, desviación estándar, Cp y Cpk por medio de datos normales, dado contrario que los datos no sean normales, se lleve a cabo la transformación y directamente calcule los valores correctos antes mencionados.



## Objetivo general

Establecer un procedimiento estadístico robusto con el fin de controlar estadísticamente un sistema productivo independientemente de la distribución probabilística a la que se ajuste la muestra extraída del proceso.

## Objetivos específicos

- Analizar las diferentes maneras de transformar datos no normales en datos normales a fin de crear una herramienta de transformación de distribución de datos.
- Establecer las bases para la creación de un sistema sencillo de transformación de datos, para así obtener resultados validos de acuerdo a las características del proceso.

## Justificación

La distribución de la producción no siempre es normal y el análisis obtenido es erróneo, siendo esto un gran problema para el proceso productivo. Por lo cual se busca una herramienta que colabore a la conversión de datos no normales a datos normales.

Beneficios:

- Mejorar el análisis de datos.
- Eliminación de falsas alarmas en el proceso.
- Reducción de costos.

## Metodología

Para verificar la normalidad o no normalidad en una muestra de datos, se retomó el trabajo realizado por (Pulido, H. y de la Vara R., 2008). Que utiliza una muestra de 100 datos no normales (de acuerdo al teorema de limite central la muestra mínima es de 30 datos). El ejemplo establece especificaciones de 1.5 y 12.5.

**Tabla 1: Muestra de datos presentados en ejemplo**

7.5	6.44	6.78	7.7	6.94	7.54
7.22	7.52	7.85	7.1	9.54	8.22
7.92	7.13	10.01	8.57	7.52	6.8



7.84	8.41	8.13	7.84	6.6	7.56
11.16	5.87	6.84	6.69	10.14	7.34
6.59	6.83	7.67	7.56	7.34	7.47
8.24	8.21	6.9	6.62	7.99	6.95
6.46	6.85	10.08	5.62	27.52	6.61
5.25	7.08	6.43	6.69	6.42	9.16
7.42	7.63	7.3	7.8	5.93	6.66
7.58	8.81	9.86	8.1	8.86	7.23
5.98	6.91	7.21	7.31	6.35	7.5
7.65	9.46	5.93	6.32	7.84	7.05
6.28	8.75	7.73	7.74	7.89	6.95
8.15	6.46	7.13	6.47	9.49	6.61
6.8	7.96	6.07	7.35	7.13	6.39
7.75	6.39	5.87	5.81		

Fuente: Gutiérrez y De la Vara, 2013.

## Análisis de información Estudio I

A continuación, se presenta un análisis de capacidad del proceso suponiendo normalidad en los datos, ver figura 1:

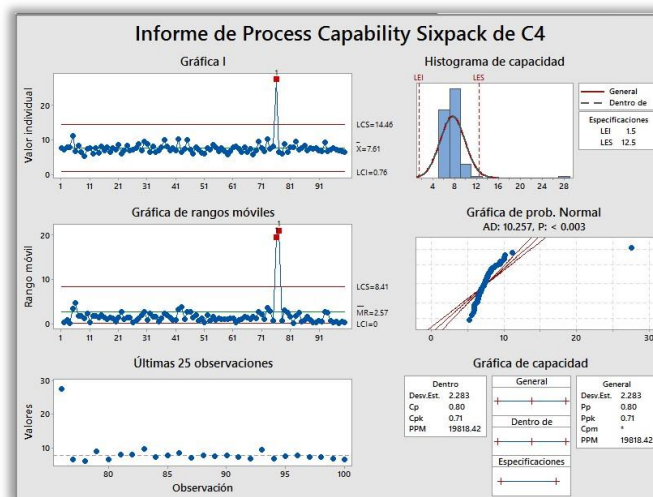


Figura 1: Análisis de capacidad del proceso suponiendo normalidad en los datos. Fuente: Propia.

## Interpretación del informe de capacidad del proceso

La Figura 1, indica que la gran mayoría de los datos se encuentran dentro de los límites de especificación, a excepción del punto número 76 el cual se puede definir



como atípico. El histograma de capacidad presenta los datos dentro de los límites de especificación, sin embargo, los datos se encuentran desfasados a la izquierda. El informe de capacidad también arrojó los siguientes parámetros, ver tabla 2.

Tabla 2: Resultado de la gráfica de capacidad. Fuente: Propia.

Gráfica de capacidad

Desv. Est.	2.283
Cp	0.8
Cpk	0.71
PPM	19818.42

Los datos que se encuentran en la tabla 2, muestran la desviación estándar, índices de capacidad ( $C_p$  y  $C_{pk}$ ) y las pérdidas por millón. El análisis realizado arrojó un  $C_p$  de 0.80, el cual indica un proceso de categoría 3 no adecuado para el trabajo y con necesidad de modificaciones urgentes. Para que el índice  $C_{pk}$  sea aceptable e indique capacidad satisfactoria, debe ser mayor a 1.25. El índice  $C_{pk}$  arrojado por el estudio fue de 0.71, lo cual indica que el proceso no tiene capacidad real o que este no está centrado. Además, se prevén pérdidas de 19818 piezas por millón producidas. En base al análisis de capacidad realizado, se decide que el proceso **no** es apto, ni cuenta con la capacidad para cumplir con las especificaciones del cliente.

#### Prueba de normalidad

Antes de analizar cualquier muestra de datos, estos deben ser sometidos a una prueba de normalidad a fin de descartar una posible no normalidad. De acuerdo con los parámetros de normalidad de la prueba Anderson-Darling, para que una serie de datos se considere normal, su valor de  $p$  debe ser mayor a 0.05. Existen otros procedimientos de prueba de normalidad como la Saphiro-wilk y kolmogorov, aunque de acuerdo con (Pulido, H. y de la Vara R., 2008), no existe una diferencia significativa en cuanto a los resultados que proporciona cada método.

En los datos proporcionados por la gráfica, claramente se puede apreciar en la parte lateral derecha un valor de  $p$  menor a 0.005, lo cual indica no normalidad. En caso de no normalidad en los datos, lo mejor es recurrir a transformarlos. En este caso se aplicó la transformación Box-Cox, la cual indica la siguiente fórmula para obtener la variable  $\lambda$ :



**Ecuación 1: Para obtener el valor de  $\lambda$ .**

$$-\ln(L(\lambda)) = -\frac{n}{2} \ln \left( \sum_{i=1}^n \frac{(x_i(\lambda) - \bar{x}_i(\lambda))^2}{x_i(\lambda)} \right) + (\lambda - 1) \sum_{i=1}^n \ln(x_i)$$

Se evaluó esta función entre valores  $[-2,2]$ , y se ubicó aquél donde la función alcanza su máximo valor óptimo:

**Tabla 3: Evaluación de  $\lambda$  para maximización de función.**

$\lambda$	-1.3	-1.2	-1.1	-1	-0.9	-0.8	-0.7
$\ln(L(\lambda))$	-0.679	-0.6248	-0.6003	-0.6059	-0.6415	-0.7074	-0.8037

Fuente: Gutiérrez y De la Vara, 2013.

El máximo valor de la función se alcanza en  $\lambda = -1.1$ . En base a la ecuación siguiente:

**Ecuación 2: Transformación de Box-Cox de acuerdo a su valor de  $\lambda$ .**

$$(\lambda) = \begin{cases} \frac{x^\lambda - 1}{\lambda} & \lambda \neq 0 \\ \ln(x) & \lambda = 0 \end{cases}$$

$\lambda$  es diferente de 0, por lo que se aplica la ecuación superior, en la que cada dato se elevará exponencialmente al valor de  $\lambda$  para así restarle 1 y esto se dividirá entre el valor de  $\lambda$ .

Además, para que el análisis de datos sea efectivo, también se deben transformar las especificaciones (sólo para el análisis de capacidad).

Luego de aplicar la ecuación, los datos transformados son los siguientes:

**Tabla 4: Datos del ejemplo transformados**

0.80999861	0.79191648	0.79836362	0.81282611	0.80116843	0.81950322
0.8057633	0.81028846	0.81484756	0.80384066	0.83304005	0.79872181
0.81576341	0.80432769	0.8369586	0.82351955	0.81028846	0.81086335
0.81471533	0.82172707	0.8184117	0.81471533	0.79503731	0.80761998



0.84509139	0.7793415	0.79943158	0.79672396	0.8379752	0.80956076
0.79484692	0.79925495	0.81241185	0.81086335	0.80761998	0.80133923
0.81974239	0.81938318	0.80048003	0.79541629	0.81666242	0.7952271
0.79231546	0.79960766	0.83750942	0.77297861	0.8853775	0.8295625
0.76239024	0.80351357	0.79171601	0.79672396	0.79151488	0.79616706
0.80882276	0.81185418	0.8070082	0.81418282	0.78078485	0.80592049
0.81114841	0.82608026	0.8357506	0.8180422	0.82659542	0.80999861
0.78196443	0.80065292	0.80560564	0.80716181	0.79008837	0.80301927
0.81213378	0.83233231	0.78078485	0.78946685	0.81471533	0.80133923
0.78862845	0.82545391	0.81323699	0.81337321	0.815373	0.7952271
0.81865645	0.79231546	0.80432769	0.79251398	0.83259918	0.79090754
0.79872181	0.81627916	0.78403629	0.80777183	0.80432769	
0.81350906	0.79090754	0.7793415	0.77786682	0.81057671	

Fuente: Propia

Tabla 5: Especificaciones reales y transformadas

Especificaciones	ES	EI
Reales	12.5	1.5
Transformación	0.85259639	0.32711242

Fuente: Propia

Después de transformar los datos, se procedió a aplicar un análisis de normalidad, para corroborar una correcta transformación en los datos:

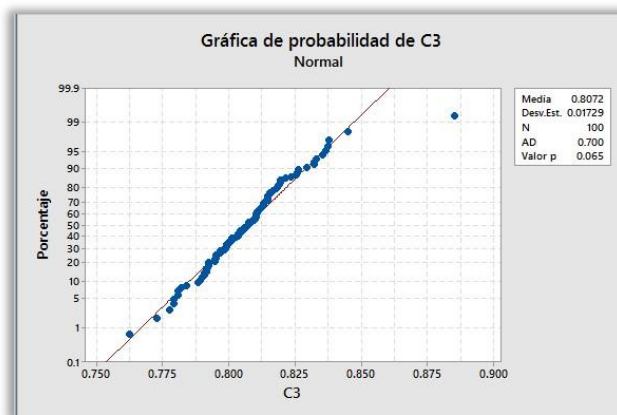


Figura 2: Prueba de normalidad a datos transformados. Fuente: Propia

En el gráfico de la figura 2, se observa que la mayoría de los datos están cerca de la media. Además, la caja de parámetros indica un valor de  $p = 0.065$ , que, de acuerdo a la prueba Anderson Darling ( $p > 0.05$ ) indica normalidad en los datos.



Luego del análisis de normalidad, se procedió a hacer un análisis de capacidad del proceso, el cual, arrojó los siguientes resultados, ver figura 3.

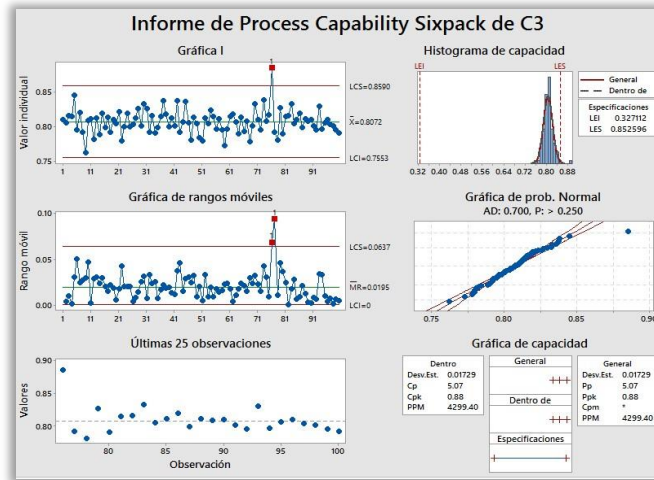


Figura 3: Análisis de capacidad a datos transformados. Fuente: Propia

Interpretación de análisis de capacidad.

En el gráfico I de la figura 3, se logra apreciar que los datos se encuentran en su mayoría, dentro de los límites de especificación, en contraste con el estudio uno, en este estudio los datos se ven mejor distribuidos y más centrados.

Tabla 6: Parámetros arrojados por análisis de capacidad de datos transformados. Fuente: Propia.

Gráfica de capacidad	
Desv. Est.	0.01729
Cp	5.07
Cpk	0.88
PPM	4299.4

Los parámetros muestran principalmente los índices de capacidad.

El índice  $C_p$  arrojó un resultado de 5.07, el cual, de acuerdo a los parámetros establecidos, se trata de un proceso de clase mundial que tiene con calidad Seis Sigma. El índice  $C_{pk}=0.88$ , de acuerdo a su interpretación Seis Sigma indica un proceso estable, sin embargo, este se puede mejorar, ya que, está desplazado a la derecha, ver figura 3 o hacia el límite superior. Sin embargo, sigue dentro de las especificaciones.

Y se prevén 4299.40 pérdidas por millón, a diferencia del estudio 1, que se pronosticaron 19818 piezas perdidas.

Vol. 6 (2019) 7º Encuentro de Jóvenes Investigadores





## Trabajo actual

En base a los estudios realizados se logró desarrollar una herramienta estadística capaz de determinar la normalidad de los datos presentados y en caso de que se trate de datos no normales, se realice la transformación de estos y así poder determinar correctamente sus índices de capacidad.

ANÁLISIS DE DATOS

libro1.xlsx    0.852598    0.327112

INSERTAR NOMBRE DE ARCHIVO    E.S.    E.I.

LEER DATOS    TRANSFORMACIÓN    5.06525

CALCULAR    VALORES DE  $\lambda$     RESULTADO Cp.

0.0172905     0    0.875861

DES. STD     1

0.807164     0.5

MEDIA     -1.1    RESULTADO Cpk.

Figura 4: Herramienta propuesta para análisis y transformación de datos. Fuente: Propia

## Resultados

En la figura 4, se muestran los datos transformados por la metodología propuesta, siendo esta una herramienta confiable, debido a que analiza la normalidad de los datos y los transforma en dado caso de que la prueba resulte no normal, con el fin calcular de forma correcta y directa los índices de Cp y Cpk, el cual, da como resultado datos veraces, por medio de un cálculo directo. Lo que darían una similitud si se hubiera realizado por medio del software Minitab, siendo no una función principal de Minitab, debido a que el análisis se lleva a cabo de forma manual para obtener los resultados como se muestra en la tabla 7 y figura 5.



**Tabla 7: Parámetros arrojados por análisis de capacidad de datos transformados. Fuente: Propia.**

Gráfica de capacidad	
Desv. Est.	0.01729
Cp	5.07
Cpk	0.88
PPM	4299.4

**Figura 5: Herramienta propuesta para análisis y transformación de datos. Fuente: Propia.**

### Trabajo a futuro

Se pretende realizar mejora continua en la herramienta estadística diseñada, ya que cuenta con algunas áreas de oportunidad, como incluir mayor rango en la variable  $\lambda$ , que automáticamente se realice la transformación adecuada a la distribución de datos con solo oprimir un botón.

### Conclusión

En base a los estudios realizados se determina que es sumamente importante analizar la normalidad de los datos recopilados en cualquier muestra de un proceso industrial antes de obtener los índices de capacidad o cualquier otro indicador. Ya que, como pudimos notar, se podría realizar una metodología errónea en la obtención de indicadores que verifiquen la calidad del proceso que se está trabajando. Un análisis erróneo conlleva a la obtención de falsas predicciones o alarmas, lo cual ocurre a menudo en la industria, resultando entre el 45% de los costos totales de la empresa, según el análisis realizado por Gutiérrez y De la Vara, mientras que para todo empresario lo más importante es la generación de utilidad.

En el ejemplo analizado se creía que la capacidad del proceso era muy mala, ya que se obtuvieron índices de capacidad que distaban mucho de la calidad Seis



Sigma e indicaban que el proceso no era apto para el trabajo, sin embargo, al efectuar el análisis y transformación correcta, se determinó que el proceso cumple con una capacidad muy buena, siendo incluso un proceso con calidad Seis Sigma. Gracias a la herramienta estadística creada, el análisis de datos resulta sumamente sencillo ya que cuenta con la transformación de Box-Cox de datos en caso de ser necesario. La transformación de Box-Cox es una prueba completa y con mayor cantidad de variables que otras, lo que la convierte en una prueba confiable y con alto grado de aceptación en cuanto a resultados obtenidos. Se planea mejorar esta herramienta estadística, a fin de ofrecer una mayor cantidad de pruebas de transformación que se adecuen al interés del usuario, incluyendo incluso otro tipo de procedimientos más robustos y así, aumentar la precisión y el nivel de confianza en los datos.

## Referencias

1. Binny, D. (septiembre de 2018). *Tomotherapy treatment site specific planning using statistical process control*. 32-39.
2. Carlos, H. P. (enero de 2016). *Aplicación del control estadístico de procesos en el control de su calidad*. *Tecnología Química*, 130- 145. Obtenido de <https://www.redalyc.org/pdf/4455/445543786011.pdf>
3. Gómez, R. C. (17 de abril de 2017). *El control estadístico de procesos*. Obtenido de <https://www.esan.edu.pe/apuntes-empresariales/2017/04/el-control-estadistico-de-procesos/>
4. Gurka, M. (2006), 'Extending the Box Cox transformation to the linear mixed model', *Journal of the Royal Statistical Society Series A*. 169, 273–288.
5. Pulido, H. y de la Vara R. (2008) Publicado: Impreso en México Título: Control estadístico de la calidad y Seis Sigma. 3<sup>ra</sup> edición. Editorial: Mc Graw Hill.
6. Salgado, D. (2017). *Pruebas de normalidad*. abril 12, 2019, de estadisticacondago Sitio web: [http://www.estadisticacondago.com/images/estadistica\\_inferencial/pruebas%20de%20normalidad.pdf](http://www.estadisticacondago.com/images/estadistica_inferencial/pruebas%20de%20normalidad.pdf)
7. Wagner, R., Haefner, B., & Lanza, G. (2018). *Function-Oriented Quality Control Strategies for High Precision Products*. *Procedia CIRP*, 57-62.
8. Xie, S., & Yang, C. (febrero de 2019). *A novel robust data reconciliation method for industrial processes*. Obtenido de <https://www.sciencedirect.com/science/article/abs/pii/S096706611830580X?via%3Dihub>