Cartas CUSUM para media y variabilidad

Cristóbal Belmar Osorio y Martín Elizondo Norambuena

Carolina Marchant Fuentes

Control Estadístico de Procesos (IES-325)

22 de Octubre, 2024

Introduccion

La carta CUSUM, conocida por sus siglas en inglés como "Cumulative Sum Control Chart", es una herramienta estadística que juega un papel fundamental en el control de calidad y la mejora continua de procesos productivos. Desarrollada por el estadístico E.S. Page en 1954 quien es un pionero en el campo del control de la calidad, donde su trabajo se centró en la aplicación de métodos estadísticos para mejorar procesos industriales, esta metodología fue creada con el propósito de superar algunas de las limitaciones sustanciales a las cartas de control tradicionales, como las cartas de control Shewhart. A diferencia de estas últimas, que detectan grandes desviaciones en un proceso, la carta CUSUM es especialmente eficaz para identificar pequeños y sutiles cambios en la media del proceso, lo que permite una mayor precisión y rapidez en la detección de problemas de calidad.

La esencia de la carta CUSUM inicia en su capacidad para acumular de manera continua las desviaciones de la media respecto a un valor objetivo o de referencia según quien lo exija. A lo largo del tiempo, estas sumas acumuladas revelan tendencias que, en otras circunstancias o en las cartas de Shewhart, mejor dicho, podrían pasar desapercibidas con métodos más simples de control de calidad. Este enfoque acumulativo permite una intervención temprana, y más bien crucial para corregir desviaciones antes de que estas afecten significativamente la calidad del producto o servicio en cuestión.

La creación de la carta CUSUM fue un hito importante en el campo del control estadístico de procesos, ya que ofreció una herramienta avanzada para la detección de cambios en los procesos productivos, especialmente aquellos que implican desviaciones pequeñas pero constantes. La sensibilidad de esta carta es vital para industrias que requieren altos estándares de calidad, o actualmente en la era 6σ, permitiendo no solo mantener el control de los procesos, sino también optimizar los sistemas de producción.

Desarrollo

¿Qué es la Carta CUSUM?

La carta CUSUM es la carta de control de sumas acumuladas, la cual incorpora directamente toda información contenida en la secuencia de los valores muéstrales, graficando las sumas acumuladas de las desviaciones que presentan los valores muéstrales respecto al valor objetivo (target).

Esta carta se puede representar como gráfico de sumas acumuladas, CUSUM tabular y en la forma de V-Mask.

Se deben cumplir los siguientes supuestos:

- Normalidad en los datos.
- Independencia de las observaciones.
- Desviación conocida o una estimación.
- Media conocida o estimación.
- Cartas de Shewart bajo control.

¿Por qué utilizar la carta CUSUM?

A diferencia de las cartas de Shewhart la cual es efectiva para corrimientos en rangos de 1.5σ a 2σ o superior la carta es efectiva, ya que son eficaces para detectar desviaciones grandes e inmediatas en el proceso, esto se representa como picos o caídas abruptas, sin embargo, no son tan sensibles a los pequeños corrimientos (cambios graduales) en la media del proceso, puesto que solo responden cuando los datos exceden estos límites. En cambio, la carta CUSUM es mucho más sensible a pequeños.

CUSUM tabular para monitorear la media del proceso

Observaciones individuales:

La CUSUM tabular funciona acumulando las desviaciones de μ_0 es decir el "objetivo" para la característica de la calidad x que se encuentran arriba del target, denotado como C^+, mientras que acumulando las desviaciones de μ_0 que se encuentran bajo el target están denotadas como C^-, estos estadísticos C, se les llama CUSUM unilaterales superior e inferior respectivamente.

$$C_i^+ = mx[0, x_i - (\mu_0 - K) + C_{i-1}^+]$$

$$C_i^- = mx[0, (\mu_0 - K) - x_i + C_{i-1}^-]$$

donde los valores iniciales $C_0^+ = C_0^- = 0$

Para calcular las CUSUM unilaterales se necesita del valor de referencia o tolerancia (K), este valor se escoge aproximadamente a la mitad del valor de μ_0 y el valor fuera de control de la media (μ_1) que es de interés para detectar con rapidez.

Si μ_1 se denota como:

$$\mu_1 = \mu_0 + \delta \cdot \sigma$$
$$\delta = \frac{|\mu 1 - \mu 2|}{\sigma}$$

Si C^+ o C^- excede el intervalo de decisión H, es decir, los límites de control de la carta CUSUM, se considera que el proceso está fuera de control, caso contrario, el proceso se encuentra en control estadístico, una de las ventajas de esta carta es su particularmente útil para determinar la causa asignable

Recomendaciones para elaborar la carta CUSUM

Se sabe que esta carta se diseña eligiendo valores para K y H los cuales se pueden escribir de la siguiente forma: - $K = k\sigma$ - $H = h\sigma$

Utilizando H = 4 o 5 y un valor K = 1/2 , se obtendrá una CUSUM que tiene propiedades convenientes de la longitud media de corrida (ARL), contra un corrimiento de aproximadamente 1σ en la media del proceso, se aprecia que si selecciona un valor $H = 4.77$, el ARL_0 de la CUSUM coincide con el valor ARL_0 para las cartas de control de Shewhart, con los límites de 3σ usuales, cabe recalcar que el ARL es el promedio de puntos antes de que la carta CUSUM detecte un cambio (corrimiento), es por ello que la carta CUSUM es sensible.

Especialistas prefieren estandarizar los valores de x_i antes de realizar los cálculos:

$$y_i = \frac{x_i - \mu_0}{\sigma}$$

Esta estandarización ofrece dos ventajas: - Diversas cartas CUSUM pueden tener los mismos valores de k y h, y las elecciones de estos parámetros no dependen de la escala, es decir, no dependen de σ. - Una CUSUM estandarizada conlleva de forma natural a una CUSUM para controlar la variabilidad.

Como se vio anteriormente, cuando se trabaja con observaciones individuales, también se puede extender al caso de promedios de subgrupos racionales $(n > 1)$, utilizando la media muestral de x_i, y σ se sustituye por $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$, como se sabe el uso de promedio en subgrupos en el caso de las cartas de Shewhart mejora sustancialmente el desempeño, sin embargo, esto no siempre ocurre con las CUSUM.

Corrimiento en la media (múltiplo de σ)	(a) CUSUM básica	(b) CUSUM-Shewhart (límites de Shewhart en 3.5σ)	(c) CUSUM con FIR	(d) CUSUM-Shewhart FIR (límites de Shewhart en 3.5σ)
0.00	465	391	430	360
0.25	139	130.9	122	113.9
0.50	38.0	37.20	28.7	28.7
0.75	17.0	16.80	11.2	11.2
1.00	10.4	9.65	5.83	6.32
1.50	5.75	5.58	3.37	3.37
2.00	4.11	3.77	2.10	2.36
2.50	3.11	2.77	1.54	1.54
3.00	2.57	2.10	1.30	1.46
4.00	2.01	1.34	1.16	1.16

CUSUM para monitorear la variabilidad del proceso

Las cartas CUSUM también pueden construirse para monitorear la variabilidad del proceso, debido a que se emplean generalmente con observaciones individualmente, pero como antes fue mencionado esta carta puede ser estandarizada, en el libro Montgomery se menciona que Hawkins sugiere crear la nueva variable estandarizada de la siguiente manera:

$$v_i = \frac{\sqrt{|y_i|} - 0.822}{0.349}$$

Hawkins sugiere que v_i son mas sensible a los corrimientos en la varianza que a los de la media, debido a que el estadístico v_i bajo control se distribuye $N(0,1)$, por esta razón es posible establecer dos CUSUM unilaterales a escala estandarizadas:

$S_i^+ = mx[0, v_i - k + S_{i-1}^+]$

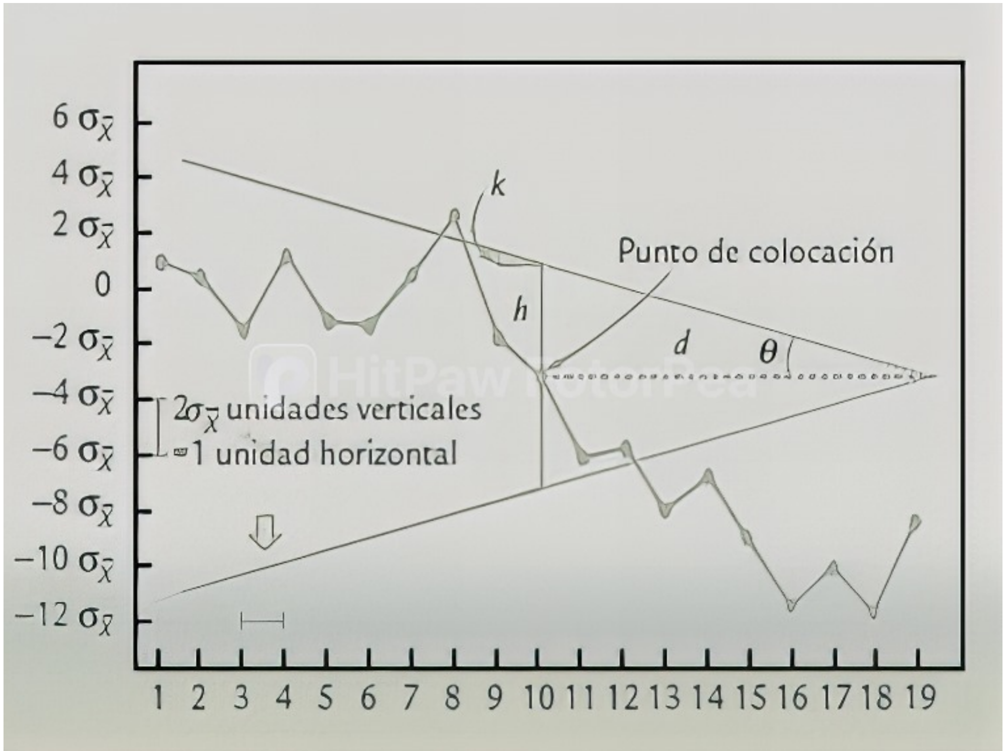
$S_i^- = mx[0, -k - v_i + S_{i-1}^-]$

Donde $S_0^+ = S_0^- = 0$ y los valores de k y h se seleccionen como en la CUSUM para controlar la media del proceso

Respecto a la interpretación, esta es similar a la interpretación de una CUSUM para la media. Si la desviación del proceso incrementa, los valores de S_i^+ se incrementarán y terminaran por exceder el h mientras que en el caso contrario los valores de S_i^- incrementaran y terminaran por exceder h

V-mask

Como antes fue dicho, para evaluar la velocidad de respuesta de la carta de control es el ARL, también conocido como la longitud promedio de corrida, es decir, es el número de puntos que en promedio se deben graficar en la carta para que esta detecte una señal de fuera de control. Existe actualmente un ARL estándar el cual es de 370.4, este nace de la división de $\frac{1}{p}$, donde p nace de la probabilidad existente en los extremos externos a más menos 3σ, siendo equivalente a 0.0027. Por lo tanto, al realizar la multiplicación se obtiene el valor estándar de ARL, por ello a partir del valor cualquier ARL que sea superior se traducirá en que va a aumentar la confiabilidad de que los datos que sean graficados están dentro de la zona de control. El valor de ARL estadísticamente significa en este caso que cada 370.4 puntos el proceso puede lanzar una falsa alarma o un error de tipo 1. En la siguiente imagen se ve expresada la CUSUM junto a la V-Mask:



Donde h es la distancia entre el punto de colocación y los brazos propios de la V-mask, mientras que k es la mitad del salto que interesa detectar, es decir, que tan sensible quiero que sea mi carta para la variable en estudio, ambos parámetros se expresan en base del error estándar, anteriormente también fue dicho que los valores típicos que se recomiendan para h y k son 5 y 0.5 respectivamente, ya que son apropiados cuando se interesa detectar un cambio de magnitud de un error estándar y un ARL de 465.

La selección de los parámetros se puede realizar con base en las siguientes formulas y tabla.

$$d = \frac{h}{k}$$

$$\theta = arctan(\frac{k}{2}) = arctan(\frac{h}{2d})$$

δ		50	100	200	300	400	500
0.25	k	0.125			0.195		0.248
	d	47.6			46.2		37.4
	L(0.25)	28.3			74.0		94.0
0.50	k	0.25	0.28	0.29	0.28	0.28	0.27
	d	17.5	18.2	21.4	24.7	27.3	29.6
	L(0.50)	15.8	19.0	24.0	26.7	29.0	30.0
0.75	k	0.375	0.375	0.375	0.375	0.375	0.375
	d	9.2	11.3	13.8	15.0	16.2	16.8
	L(0.75)	8.9	11.0	13.4	14.5	15.7	16.5
1.0	k	0.50	0.50	0.50	0.50	0.50	0.50
	d	5.7	6.9	8.2	9.0	9.6	10.0
	L(1.0)	6.1	7.4	8.7	9.4	10.0	10.5
1.5	k	0.75	0.75	0.75	0.75	0.75	0.75

δ		50	100	200	300	400	500
	d	2.7	3.3	3.9	4.3	4.3	4.7
	L(1.5)	3.4	4.0	4.6	5.0	5.2	5.4
2.0	k	1.0	1.0	1.0	1.0	1.0	1.0
	d	1.5	1.9	2.2	2.4	2.5	2.7
	L(2.0)	2.26	2.63	2.96	3.15	3.3	3.4

δ = Brinco a detectar en unidades del error estándar.

$$\delta = \frac{\text{Cambio a detectar}}{\sigma_{\bar{x}}}$$

Cálculo de las sumas acumuladas:

$$S_m = (\overline{X}_1 - \mu) + (\overline{X}_2 - \mu) + (\overline{X}_3 - \mu) + \cdots + (\overline{X}_m - \mu) = \sum_{i=1}^{m} (\overline{X}_i - \mu)$$

Ejercicio V-mask

Una máquina automática llena paquetes de harina con un peso nominal de 80 onzas. Interesa monitorear esta máquina para detectar cambios pequeños de magnitud 0.15 onzas. Por medio de información histórica se sabe que la media y la desviación estándar del peso de los paquetes es de 80.00 y 0.2 onzas, respectivamente. Cada media hora se sacan en forma aleatoria cuatro paquetes y se pesan. Las medias de las últimas 20 muestras, los rangos y las sumas que acumulan la desviación de las medias respecto a la media histórica (80.00).

Observaciones

Muestra	Media	Rango	Suma acumulada
1	79.90	0.35	-0.10
2	79.91	0.43	-0.19
3	79.89	0.63	-0.30
4	80.05	0.50	-0.25
5	79.94	0.50	-0.31
6	79.95	0.18	-0.36
7	79.88	0.45	-0.48
8	79.96	0.36	-0.52
9	80.27	0.65	-0.25
10	79.87	0.23	-0.38
11	79.87	0.50	-0.51
12	80.04	0.23	-0.47
13	80.04	0.56	-0.43
14	80.04	0.15	-0.39
15	80.23	0.33	-0.16
16	80.23	0.62	0.07
17	80.28	0.34	0.35
18	80.13	0.33	0.48
19	80.05	0.45	0.53
20	80.15	0.47	0.68

Datos:

- $N = 80$ onzas
- Cambio a detectar: 0.15 onzas
- Media histórica (μ) = 80
- Desviación estándar (σ) = 0.2

Solución

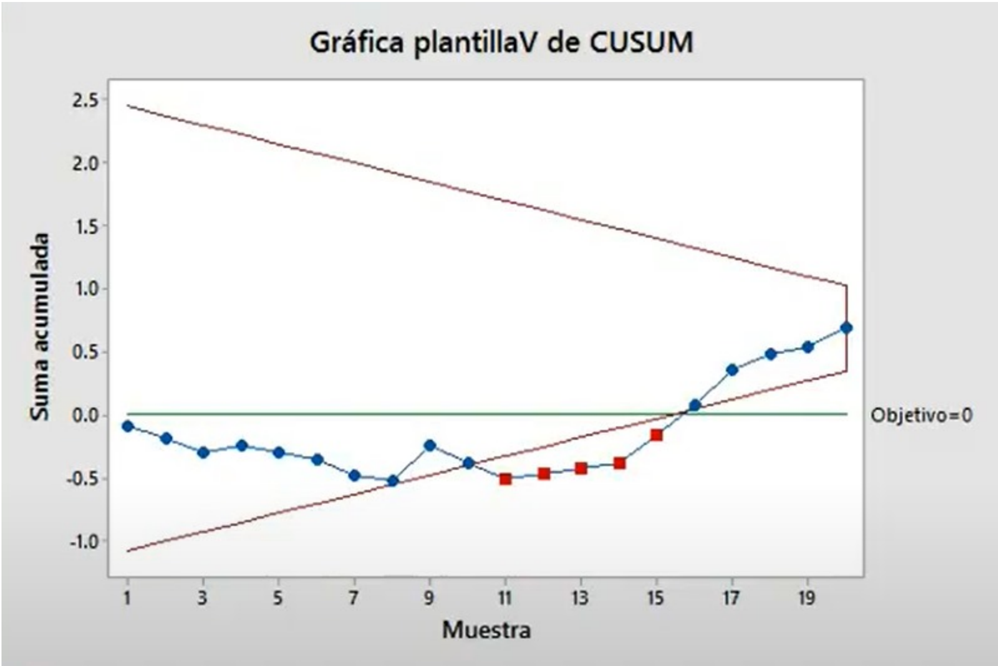
- Cálculo de valores:

$$\sigma_x = \frac{0.2}{\sqrt{4}} = 0.1$$

$$\delta = \frac{0.15}{0.1} = 1.5$$

$$\theta = \arctan\left(\frac{k}{2}\right) = \arctan(0.375) = 20.6°$$

- Con $\delta = 1.5$ y $ARL_0 = 400$, los parámetros son:
- $d = 4.5$
- $k = 0.75$
- $h = (0.75)(4.5) = 3.37$



Aplicación de las cartas CUSUM

Las cartas CUSUM se aplican en sectores que requieren monitoreo continuo de procesos para asegurar la calidad y consistencia de productos o servicios. En manufactura, permiten controlar variables como peso o espesor, detectando desviaciones tempranas y reduciendo costos.

- En la industria alimentaria controlan parámetros de seguridad como acidez o temperatura.
- En la industria farmacéutica, aseguran la dosificación adecuada y el cumplimiento de regulaciones.

- En ingeniería y construcción, monitorizan la calidad de materiales como el acero. En servicios y finanzas, controlan tiempos de respuesta y riesgos financieros.
- También se usan en hospitales para monitorear procesos médicos, y en el sector aeroespacial para el control de calidad de componentes clave.

En resumen, las CUSUM son esenciales donde el control continuo es vital para la calidad y seguridad.

Ejercicios

Datos simulados

Monitoreo del rendimiento de cultivos de trigo en una región agrícola:

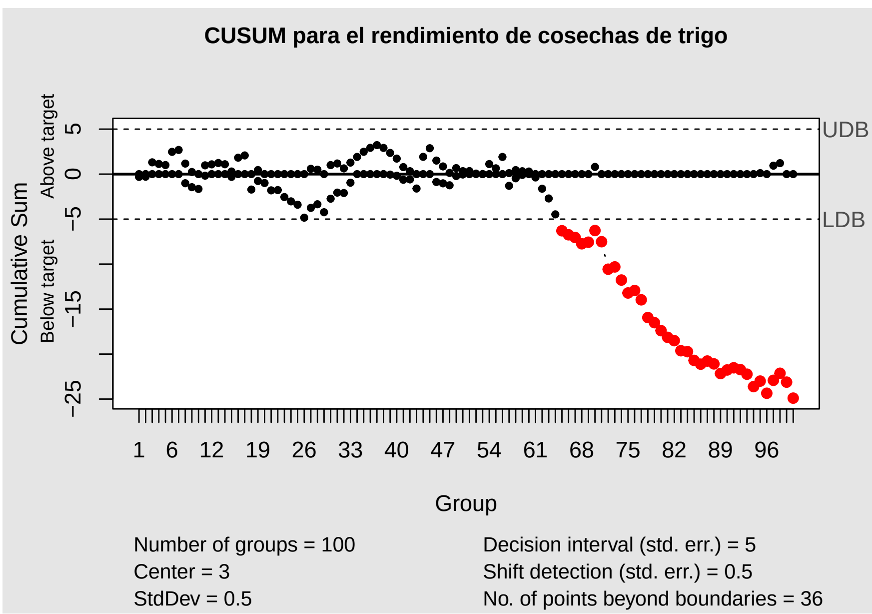
La región del Maule depende en gran medida del cultivo de trigo como fuente principal de ingreso. Se ha observado que los rendimientos de las cosechas han sido consistentes durante los últimos años. Sin embargo, debido a factores como cambios en las condiciones climáticas y la adopción de nuevas prácticas de fertilización, los agricultores están preocupados de que los rendimientos puedan estar fluctuando más de lo habitual.

Para identificar posibles variaciones en el rendimiento del trigo por hectárea, se desea implementar una carta CUSUM que permita detectar cambios pequeños, pero significativos en el rendimiento promedio, para que los agricultores puedan ajustar rápidamente sus prácticas agrícolas. Los datos históricos indican que el rendimiento promedio ha sido de 3 toneladas por hectárea con una desviación estándar de 0.5 toneladas.

- X : Rendimiento del trigo de 100 semanas por hectárea en tonelada.
- H_0 : El proceso está bajo control estadístico.
- H_1 : El proceso no está bajo control estadístico.

Desarrollo

Se crean los datos y se crea la carta de control con la librería "*qcc*".



Durante las primeras 60 semanas, los puntos se encuentran cerca de la línea central, lo que indica que el proceso se mantuvo dentro del rendimiento esperado, con pequeñas fluctuaciones normales. No hay señales de alarmas importantes ni de desviaciones significativas respecto al objetivo de 3 toneladas. A partir de la semana 61 observamos un claro descenso de los puntos hacia la parte inferior del gráfico, lo que indica una tendencia hacia un rendimiento más bajo, lo que significa que el proceso está fuera de control con un nivel de significación del 0.0027.

Variabilidad en la Producción de Resistencias

En una fabricación de componentes eléctricos, es crucial asegurar que los productos finales cumplan con las especificaciones técnicas requeridas, en particular con el valor de resistencia eléctrica, que en este caso es de 100 ohmios. Además de controlar que la media de las resistencias fabricadas sea consistente con el objetivo, es fundamental monitorear la variabilidad en el proceso de producción, es decir, las fluctuaciones en la desviación estándar de las resistencias. Por esta razón se desea monitorear la variabilidad en las resistencias de 50 componentes eléctricos.

Definición de variable:

- X : Resistencia eléctrica de los componentes en ohmios
- H_0 : El proceso esta bajo control estadístico
- H_1 : El proceso no esta bajo control estadístico
- Sea $\alpha = 0.0027$

Desarrollo

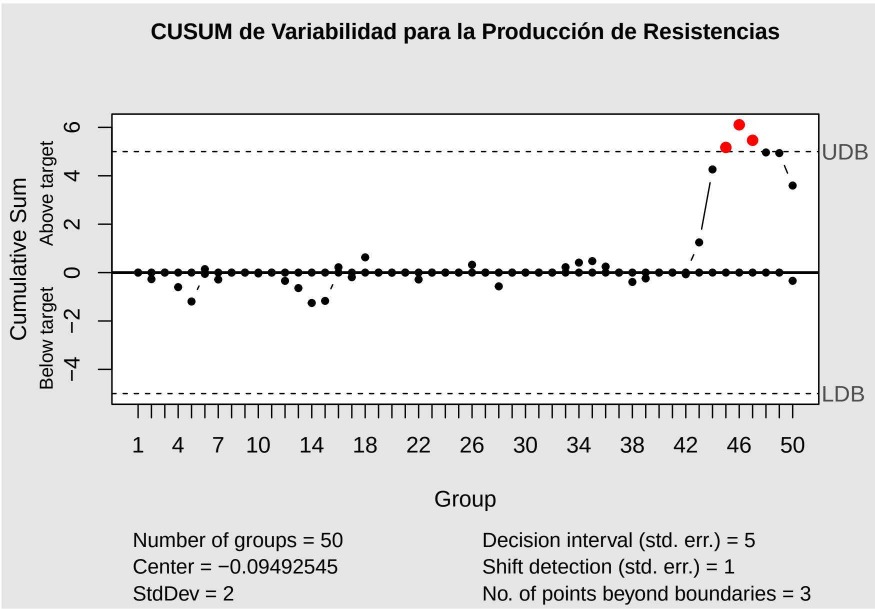
- Paso 1: Estandarizar la variable Para estandarizar la variable se aplica la siguiente fÓrmula:

$$y = \frac{(x - \mu)}{\sigma}$$

- Paso 2: estandarizar nuevamente utilizando $\mu = 0.822$ y $\sigma = 0.349$

$$v = \frac{(\sqrt{|y|} - 0.822)}{0.349}$$

Posterior a la estandarizacion utilizamos la librería "*qcc*" y calculamos la CUSUM para la variabilidad.



Al analizar la carta se aprecia que durante las primeras 30 observaciones, el proceso parece estar bajo control en términos de variabilidad, ya que los puntos se mantienen dentro de los límites de decisión. Sin embargo, a partir de la observación 40, la carta CUSUM comienza a mostrar puntos fuera de control, lo que indica un aumento en la variabilidad. Por lo que el proceso está fuera de control estadístico con un nivel de significación del 0.0027 Es recomendable investigar las posibles causas de este aumento en la variabilidad para corregir el proceso si es necesario.

Datos reales

Para realizar un ejercicio con datos reales se decidió utilizar una base de datos de la página *Kaggle* de la cual se seleccionó la base de datos "*Wine Quality*" la cual posee datos de la preparación de vino con uvas rojas y con uvas blancas, en estos ejemplos se optó por utilizar la uva roja y monitorear el *pH*.

Control de calidad del pH en la producción de vino de uvas rojas

El pH del vino es una medida crítica en la industria vinícola, ya que afecta el sabor, la estabilidad microbiológica y la longevidad del vino. Por esta razón un productor desea monitorear el pH del vino tinto a lo largo del tiempo para asegurar que se mantenga dentro de los rangos de calidad establecidos (pH entre 3.2 y 3.6). Un pH fuera de estos límites podría indicar problemas con la fermentación o con la acidez, lo que podría afectar la calidad del vino. Para esto se tomó una muestra de 380 vinos y se midió su pH.

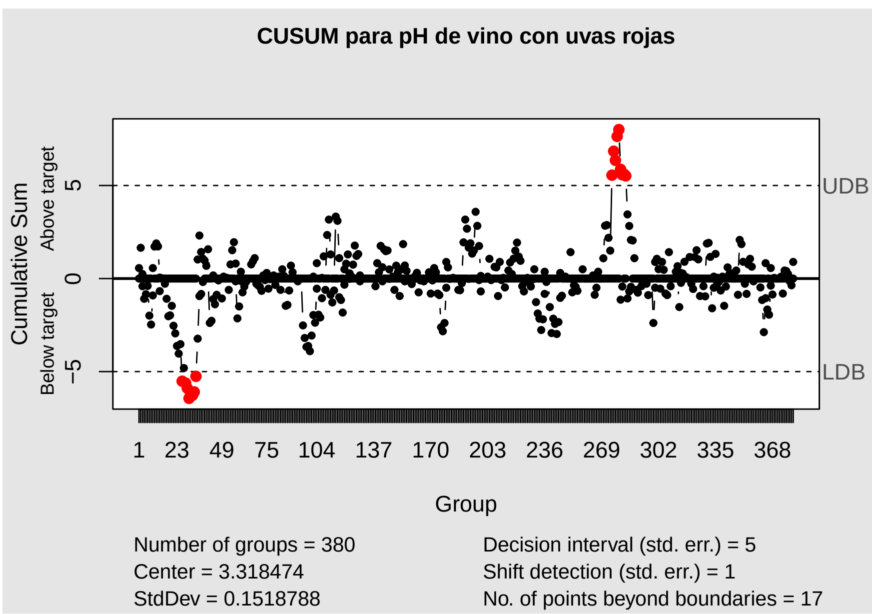
- X : Nivel de PH de 380 muestras de vino tinto.
- H_0 : El proceso está bajo control estadístico
- H_1 : El proceso no esta bajo control estadístico
- Sea $\alpha = 0.0027$

Comprobar normalidad

Para comprobar la normalidad de los datos se aplicó el test de Anderson Darling

```
Al aplicar el test de Anderson Darling entrega un valor p igual a 0.07823849
 por lo que los datos siguen una distribucion normal
```

Al confirmar el supuesto de normalidad y suponiendo que el proceso anteriormente estaba bajo control estadístico, creamos la carta CUSUM con la librería "*qcc*".



Al analizar la carta CUSUM muestra que el proceso estuvo bajo control en la mayoría de las observaciones, pero hubo ciertos momentos en los que se identificaron posibles problemas o variaciones significativas. Estos momentos fuera de control se deben investigar para identificar y corregir las causas que llevaron a esas fluctuaciones. Es posible que las intervenciones o ajustes sean necesarios para mantener el proceso dentro de los límites controlados. Estos cambios se pueden deber a las condiciones ambientales como fluctuaciones en las condiciones de almacenamiento, temperatura o en el proceso de fermentación del vino.

Control de variabilidad del pH en la producción de vino de uvas rojas

En la producción de vino, es fundamental que el pH se mantenga dentro de un rango controlado para asegurar la calidad del producto. No solo es importante controlar el valor promedio del pH, sino también la variabilidad del mismo, ya que fluctuaciones excesivas en la variabilidad pueden afectar negativamente el sabor y la estabilidad del vino. Utilizando los datos del ejercicio anterior se calculará la CUSUM para la variabilidad.

Desarrollo

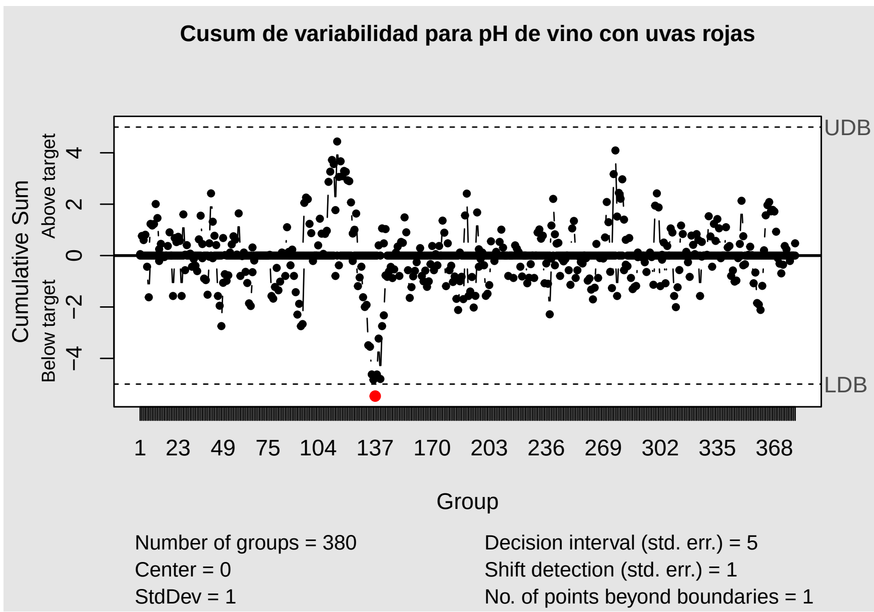
- Paso 1: Estandarizar la variable Para estandarizar la variable se aplica la siguiente fórmula:

$$y = \frac{(x - \mu)}{\sigma}$$

- Paso 2: estandarizar nuevamente utilizando $\mu = 0.822$ y $\sigma = 0.349$

$$v = \frac{(\sqrt{|y|} - 0.822)}{0.349}$$

Posterior a la estandarización utilizamos la librería "*qcc*" y calculamos la CUSUM para la variabilidad



Al analizar la variabilidad del pH del vino de uvas rojas está en gran medida bajo control, con solo un punto fuera de los límites de control inferior. Se recomienda investigar qué sucedió alrededor del grupo 137 para comprender la causa de la variabilidad inusualmente baja. Esto podría deberse a un cambio en las condiciones de producción o a una fluctuación en la calidad de las uvas utilizadas. Además de esto no hay otros indicios de que la variabilidad esté fuera de control, lo que es un buen signo de estabilidad en el proceso de producción del vino.

Referencias

Montgomery, D. C. (2005). Control estadístico de la calidad (3.ª ed.). Limusa Wiley.

Page, E. S. (1954). Continuous inspection schemes. Biometrika, 41(1-2), 100-115. https://doi.org/10.1093/biomet/41.1-2.100

Mishra, A., & Waghmare, V. (2021). A CUSUM chart using absolute sample values to monitor process mean and variability. IEEE Access, 9, 142040-142048. https://doi.org/10.1109/ACCESS.2021.3116634

Mora D, R. (2017, 20 de febrero). Cartas CUSUM bilateral y unilateral. Teoría y ejemplo en Minitab [Video]. YouTube.